

Bedingungen für die Äquivalenz
von Papier-Bleistift-Version und
Computerversion bei Leistungstests

Das Promotionsgesuch wurde eingereicht am 29. Januar 2010.

Die Arbeit wurde angeleitet von Prof. Dr. Helmut Lukesch.

Prüfungsausschuss:

Vorsitzender: Prof. Dr. Klaus-Peter Wild

Erstgutachter: Prof. Dr. Helmut Lukesch

Zweitgutachter: Prof. Dr. Karl-Heinz Bäuml

Bedingungen für die Äquivalenz von Papier-Bleistift-Version und Computerversion bei Leistungstests

Inaugural-Dissertation zur Erlangung der Doktorwürde
der Philosophischen Fakultät II
(Psychologie, Pädagogik und Sportwissenschaft)
der Universität Regensburg
vorgelegt von
Ludwig Kreuzpointner
aus Regensburg
2010

Universitätsverlag Regensburg

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

1. Auflage 2010

© 2010 Universitätsverlag, Regensburg

Leibnizstraße 13, 93055 Regensburg

Umschlaggestaltung: Ludwig Kreuzpointner

Layout: Ludwig Kreuzpointner

Druck: Docupoint, Madgeburg

ISBN 978-3-86845-059-0

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung des Verlags ist es nicht gestattet, dieses Buch oder Teile daraus auf fototechnischem oder elektronischem Weg zu vervielfältigen.

Weitere Informationen zum Verlagsprogramm erhalten Sie unter:

www.universitaetsverlag-regensburg.de

Inhaltsverzeichnis

1.	Einleitung.....	9
2.	Zur Faktorenstruktur des Leistungsprüfsystems (LPS) von Horn (1962; 1983) ..	11
2.1	Faktorenstruktur des LPS in einzelnen Untersuchungen	11
2.2	Reanalyse einzelner Untersuchungen	17
2.3	Korrelative Zusammenhänge	22
2.3.1	Korrelationen mit anderen Tests	22
2.3.2	Korrelationen mit schulischen Leistungsmaßen	25
2.4	Diskussion	28
2.5	Schlussfolgerung und Revisionsvorschlag	29
3.	Das neue Leistungsprüfsystem (LPS-neu)	33
3.1	Die Subtests des revidierten Leistungsprüfsystems	33
3.2	Das Intelligenzmodell des LPS-neu	37
4.	Äquivalenz von Papier-Bleistift-Version und Computerversion	43
4.1	Äquivalenzstudien	43
4.2	Zur Äquivalenz des LPS	45
4.3	Bedingungen für Äquivalenz	47
4.3.1	Psychometrische Äquivalenz	47
4.3.2	Experimentelle Äquivalenz	48
4.3.3	Relativität der Äquivalenz	48
4.4	Institutionelle Vorgaben	48
5.	Die Computerversion des neuen LPS	51
5.1	Bedingungen der Computerversion	51
5.2	Ableitung von Hypothesen auf Grundlage der Bedingungen der Computerversion des LPS	53
6.	Inferenzstatistische Methoden	55
6.1	Mittelwertsvergleiche	55
6.2	Schiefe und Kurtosis	56
7.	Verfahren zur Überprüfung der Äquivalenz	59
7.1	Unterschiedlichkeit der Bedingungen der CV	59
7.2	Unterschiedlichkeit von PBV und CV	62
7.3	Vergleich der Varianzen	65
7.4	Analyse der Zusammenhänge	66
7.5	Gleichzeitiger Vergleich von Mittelwert, Varianz und Kovarianz	66
8.	Testtheoretische Grundlagen	69
8.1	True Score	69

8.2	Paralleltests.....	70
8.3	Attenuitäts-Korrektur (correction for attenuation).....	71
8.4	Testzerlegung.....	71
8.5	Reliabilitäts-Validitäts-Optimierungsparadoxon	72
8.6	Itemanalyse	72
9.	Untersuchung 1: Vortestung an Studenten	75
9.1	Papier-Bleistift-Testung der studentischen Stichprobe	77
9.2	Computer-Testung der studentischen Stichprobe.....	78
9.3	Vergleich der Durchführungsmodi und Konsequenzen der Vorstudie	80
10.	Untersuchung 2: Testung an Berufsschülern	87
10.1	Design	87
10.2	Psychometrische Güte der Papier-Bleistift-Version.....	90
10.3	Psychometrische Güte der Computer-Version	96
10.4	Vergleich der psychometrischen Indizes von PBV und CV zu den beiden Erhebungszeitpunkten	103
10.5	Vergleich der Mittelwerte der Durchführungsbedingungen der CV	104
10.6	Vergleich der Mittelwerte der PBV und der CV	113
10.7	Vergleich der Korrelationen von PBV und CV.....	132
10.8	Vergleich der Varianzen.....	135
10.9	Schlussfolgerungen für die einzelnen Subtests	136
11.	Die Struktur des neuen Leistungsprüfsystems	141
11.1	Datenaufbereitung.....	141
11.2	Modellschätzung	143
11.3	Modellgüte	144
11.4	Modellvergleich	148
11.5	Struktur des LPS-neu in der Computerversion	150
12.	Resümee	153
12.1	Unberücksichtigte Aspekte	153
12.1.1	Der Einfluss des Geschlechts auf die Testleistung.....	153
12.1.2	Der Zusammenhang der Testleistung mit der Schulleistung	153
12.1.3	Probabilistische Testtheorie.....	154
12.2	Zur Revision des LPS und Anwendbarkeit für die Fragestellung	154
12.2.1	Eigenschaften der Papier-Bleistift-Version	155
12.2.2	Eigenschaften der Computerversion.....	155
12.3	Bedingungen der Äquivalenz.....	156
12.3.1	Präsentationsform	156
12.3.2	Antwortformat.....	157

12.3.3	Zeitvorgabe	159
12.3.4	Modus und Testzeitpunkt	160
12.3.5	Fazit.....	161
13.	Zusammenfassung	163
14.	Literaturverzeichnis.....	165
15.	Anhang.....	173
15.1	Testbögen des neuen LPS.....	173
15.2	Instruktion des neuen LPS	177
15.3	Verteilungen der Papier-Bleistift Version der Berufsschüler-SP	183
15.4	Verteilungen der Papier-Bleistift Version der Berufsschüler-SP getrennt nach PBV zu T1 und PBV zu T2	185
15.5	Itemkennwerte im Kontext der KTT.....	187
15.5.1	Itemkennwerte Subtest 1.....	188
15.5.2	Itemkennwerte Subtest 2.....	189
15.5.3	Itemkennwerte Subtest 3.....	190
15.5.4	Itemkennwerte Subtest 4.....	191
15.5.5	Itemkennwerte Subtest 5.....	192
15.5.6	Itemkennwerte Subtest 6.....	193
15.5.7	Itemkennwerte Subtest 7.....	194
15.5.8	Itemkennwerte Subtest 8.....	195
15.5.9	Itemkennwerte Subtest 9.....	196
15.5.10	Itemkennwerte Subtest 10	197
15.5.11	Itemkennwerte Subtest 11	198
15.6	ANOVA Tabellen der Computerbedingungen	199
15.6.1	Subtest 1.....	199
15.6.2	Subtest 2.....	199
15.6.3	Subtest 3.....	200
15.6.4	Subtest 4.....	200
15.6.5	Subtest 5.....	200
15.6.6	Subtest 6.....	201
15.6.7	Subtest 7.....	201
15.6.8	Subtest 8.....	201
15.6.9	Subtest 9.....	202
15.6.10	Subtest 10	202
15.6.11	Subtest 11	202
15.7	Mittelwerte der Rohwertsummen	203
15.8	ANOVA Tabellen der Äquivalenzüberprüfung.....	210

15.8.1	Subtest 1.....	210
15.8.2	Subtest 2.....	210
15.8.3	Subtest 3.....	211
15.8.4	Subtest 4.....	211
15.8.5	Subtest 5.....	211
15.8.6	Subtest 6.....	212
15.8.7	Subtest 7.....	212
15.8.8	Subtest 8.....	212
15.8.9	Subtest 9.....	213
15.8.10	Subtest 10	213
15.8.11	Subtest 11	213

1. Einleitung

In der Testdatenbank PSYINDEX des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID) werden nach Angaben der Homepage über 5500 Testnachweise von Tests, die nach 1945 veröffentlicht wurden, geführt; jährlich kommen ca. 200 Verfahren hinzu¹. Im Hogrefe Test System (HTS) werden bereits über 300 Testverfahren als Computerversion (CV) angeboten². Für viele dieser Verfahren steht der Nachweis, dass die Ergebnisse der CV zu den Ergebnissen der Papier-Bleistift-Version (PBV) äquivalent sind, noch aus. Auch wenn für das HTS – ebenso wie bei vermutlich den meisten anderen Testsystemen – über die anonyme Rücksendung der Testergebnisse das Ziel verfolgt wird, zusätzliche Normierungen der CV zu erstellen, kann dies nur als Zwischenlösung gesehen werden. Für die Konstruktion von computerbasierten Tests auf Grundlage bestehender Verfahren gibt es jedoch klar formulierte Vorgaben der International Test Commission (ITC, 2005; siehe Kapitel 4.4). Hierbei wird insbesondere die Überprüfung der Äquivalenz der Testverfahren als PBV und CV gefordert.

Eine bereits bestehende PBV eines Tests kann in vielfältiger Form am Computer dargeboten werden. Die Grundthese vorliegender Arbeit ist die Annahme, dass die Äquivalenz, also die Vergleichbarkeit der Ergebnisse zweier Testversionen, abhängig von der gewählten Form der Übertragung des Tests auf den Computer ist. Wenn es möglich ist, Bedingungen für die Äquivalenz zu identifizieren, kann somit die Chance, dass bei der Konstruktion einer CV eine der PBV äquivalente Form gefunden wird, bereits im Vorfeld der Äquivalenzuntersuchung erhöht werden.

Um dieser Fragestellung nachzugehen, wird das Leistungsprüfsystem von Horn (LPS; 1962, 1983) überarbeitet und in unterschiedlichen Versionen am Computer dargeboten. Es werden die Darstellung der Items auf dem Bildschirm, die Angabe der Bearbeitungszeit und die Korrekturmöglichkeiten bereits bearbeiteter Items systematisch variiert, so dass für die Äquivalenz von PBV und CV günstige Bedingungen identifiziert werden, die bei der Übersetzung anderer Tests auf den Computer Berücksichtigung finden können.

Im Folgenden werden nun zuerst das LPS und der Revisionsvorschlag vorgestellt (Kapitel 2 und 3). Anschließend werden das Konzept der Äquivalenz (Kapitel 4) und darauf aufbauend die Konzipierung der Bedingungen der Computerversion des LPS-neu (Kapitel 5) dargestellt. Neben den zur Überprüfung der Äquivalenz relevanten Methoden (Kapitel 7) wird auf einige in diesem Kontext interessante Aspekte der klassischen Testtheorie (Kapitel 8) und Kenngrößen der Testkonstruktion (Kapitel 6) detaillierter eingegangen. Zur Überprüfung des Designs wurde zunächst eine studentische Stichprobe (Kapitel 9) untersucht, bevor die Äquivalenz des LPS-neu an einer größeren Zahl Berufsschüler überprüft wurde (Kapitel 10). Anhand der erhobenen Daten wird die Struktur des LPS-neu konfirmatorisch getestet (Kapitel 11). Die Implikationen und Schlussfolgerungen werden in Kapitel 12 diskutiert.

1 <http://www.zpid.de/index.php?wahl=products&uwahl=fee&uuwahl=psyndexinfo> [Stand: 02.07.2010]

2 <http://www.unifr.ch/ztd/HTS/infest/WEB-Informationssystem/de/4de001/webcatalog.htm> [Stand: 02.07.2010]

2. Zur Faktorenstruktur des Leistungsprüfsystems (LPS) von Horn (1962; 1983)

Es existieren bereits Veröffentlichungen, die sich mit der Faktorenstruktur des LPS beschäftigen (Langfeldt, 1975; Steck, Rausche & Steege, 1975; Steller & Stürmer, 1984). Deren Ergebnisse sollen aber in dieser Arbeit nicht exklusiv, sondern zusammen mit den Ergebnissen anderer Studien, in deren Rahmen Daten zum LPS in angemessener Vollständigkeit veröffentlicht wurden, vorgestellt werden. Acht Studien (17 Stichproben) enthalten eine Interkorrelationsmatrix der Subtests, so dass eine erneute Faktorisierung mittels gleicher Verfahren und Kriterien berechnet werden kann. Diese soll im Anschluss an die Ergebnisse, die in den Arbeiten präsentiert werden, dargestellt werden.

2.1 Faktorenstruktur des LPS in einzelnen Untersuchungen

Bittner³ (1960) untersuchte 1003 Schüler im Alter zwischen 13 und 20 Jahren mittels einer Vorform des LPS (Horn, 1959, zit. nach Bittner, 1960). Diese Form bestand aus 13 Subtest, deren Items in allen außer einem – Subtest 7 der Vorform hatte eine andere Darstellung, entspricht aber prinzipiell Subtest 8 der aktuellen Form – der Systematik des aktuellen LPS (Horn, 1983) entsprachen. Subtest 13 und 14 waren noch nicht enthalten. Wegen des Vorliegens sowohl anderer Items als auch einer anderen Reihenfolge der Items werden die Daten aufgrund der Größe der Stichprobe zwar präsentiert, aber nur unter Vorbehalt interpretiert. Es wurden zwei Gruppen auf Grundlage des oberen und des unteren Quartils eines Neurotizismusfragebogens gebildet und verglichen. Die Gruppe der „Labilen“ bestand aus 259 Schülern ($M = 15;6$ Jahre), die „Stabilen“ bildeten 262 Schüler ($M = 15;7$ Jahre). Die Faktorenanalyse hielt sich an die Kriterien, die Thurstone (1947) formulierte (Zentroidmethode mit geometrischer Rotation), und kam nach mehreren Rotationen zu dem Ergebnis, dass bei den „Labilen“ drei und bei den „Stabilen“ sechs bzw. fünf Faktoren extrahiert werden können (vgl. Tabelle 2.1).

Tabelle 2.1: Interpretationen der Faktorenanalyse der Untersuchung von Bittner (1960), Zuordnung der Skalen zu den Faktoren nach mehreren geometrischen Rotationen

	Labile ($n = 259$)	Stabile ($n = 262$)
I	1 ¹ , 2	1, 2, 5, 6, 12
II	3, 7, 9, 10, 11, ($Sz \sim 8$)	3, 4, 10
III	4, 5, 6, 12, 15	7, ($Sz \sim 8$), 9
IV		11
V		15
VI		(Reasoning beeinflusst durch Schulwissen) ²

Bemerkung: ¹Die Subtests wurden in der Vorform in einer anderen Reihenfolge vorgegeben. Zur leichteren Vergleichbarkeit wird in der Tabelle die Nummerierung der aktuellen Version verwendet.

²Der VI. Faktor bestünde aus Subtests, die höher auf anderen Faktoren laden, wird aber dennoch von Bittner (1960) interpretiert.

³ Die Untersuchung von Lienert (1966) basiert auf den Daten von Bittner (1960). Die Darstellung der Daten ist bei Lienert allerdings verwirrend und die Beschreibungen so, als ob das eigentliche LPS benutzt worden wäre. Zudem stimmen offensichtlich Tabellen und deren Beschreibung nicht überein.

Brandstätter (1972) untersuchte 271 Schüler an den sechs Höheren Schulen der Stadt Essen. Das mittlere Alter betrug 17;6 Jahre ($SD = 3;3$ Jahre). Zusätzlich zum LPS wurden Persönlichkeitsfragebögen vorgelegt. Als Kriterium für die Zahl der Faktoren legte Brandstätter die Zahl nicht-negativer Eigenwerte der Korrelationsmatrix fest, so dass sieben nach Varimax rotierte Faktoren resultierten (Tabelle 2.2).

Tabelle 2.2: Interpretation der Faktorenanalyse der Untersuchung von Brandstätter (1972, $n = 271$)

	Interpretation	Subtests
I	Perzeptiv-analytische Strukturierung bei figuralen und räumlichen Konfigurationen	8, 9, 10
II	Verbalfaktor	1+2 ¹ , 5, 12
III	Geschwindigkeits- und Mengenleistung bei einfachen perzeptiven und numerischen Aufgaben	13, 14, 15
IV	Induktiv-schlussfolgerndes Denken	3, 4
V	Geschwindigkeit und Gestaltbildung (Bilderkennen)	11
VI	Flüssigkeit (Wortfinden)	6
VII	Veranschaulichung räumlicher Verschiebungen (Anagramme und Spiegelbilder)	7

Bemerkung: ¹Die Subtests 1 und 2 wurden zusammengefasst betrachtet.

Greif (1972) untersuchte neben dem LPS noch andere Gruppenintelligenztests. Diese Tests legte er 302 in der Ausbildung befindlichen Bereitschaftspolizisten vor ($M = 22;4$ Jahre, $SD = 5;10$ Jahre). Die fünf extrahierten und nach Varimax rotierten Faktoren genügen sowohl der Kaiser-Guttman-Regel mit Eigenwerten größer 1, als auch dem Scree-Test und dem Kriterium, dass ein Faktor mehr als 5 % der Varianz klären soll (Tabelle 2.3).

Tabelle 2.3: Interpretation der Faktorenanalyse der Untersuchung von Greif (1972, $n = 302$)

	Interpretation	Subtests
I	Räumliches Vorstellungsvermögen	3, 4, 7, 8, 9, 10
II	Bemerken von Fehlern, Heraussuchen und Zählen von Zahlen, Wahrnehmungstempo	13
III	Sprachgebundenes Denken	1+2, 5, 12
IV	Tempo bei einfachen Aufgaben	14, 15
V	Einfallsgeschwindigkeit	6, 11

Kury (1972) legte LPS und IST 187 Psychologiestudenten des 1. und 2. Semesters vor ($M = 22;2$ Jahre, $SD = n. a.$) vor. Es werden mehrere Kriterien für die Zahl der Faktoren angegeben (Scree-Test, 5 % Varianzklärung, zwei nicht weiter beschriebene Kriterien von Tucker und Larvel & Maxwell sowie das Bestreben, möglichst keine Faktoren mit nur einem Subtest und pro Subtest nur eine essentielle Ladung auf einem Faktor zu erreichen). Welches Kriterium nun zum Einsatz kam, ist nicht ersichtlich. Die nach Varimax rotierte Faktorenlösung ergibt vier Faktoren (Tabelle 2.4), wobei Subtest 5 und 12 nicht aufgenommen wurden, da ihre Ladungen zu gering ($< .5$) gewesen seien.

Tabelle 2.4: Interpretation der Faktorenanalyse der Untersuchung von Kury (1972, $n = 187$)

	Interpretation	Subtests
I	Räumliches Vorstellungsvermögen	3, 4, 8, 9, 10
II	Allgemeinbildung und Rechtschreibkenntnisse	1, 2
III	Initiative und Entschlusskraft	(5,) 6, 13
IV	Wahrnehmungstempo und Vorstellungskraft	7, 11, (12,) 14

Langfeldt (1975) testete 106 Studenten der Sonderpädagogik ($X_{med} = 26$ Jahre, $Sp = [22, 35]$) mit dem LPS. Nur zwei Faktoren hatten einen Eigenwert größer 1 (Tabelle 2.5). Weitere Kriterien und die Art der Rotation zur Lösung sind nicht angegeben.

Tabelle 2.5: Interpretation der Faktorenanalyse der Untersuchung von Langfeldt (1975), rotierte Lösung von Sonderpädagogikstudenten ($n = 106$)

	Interpretation	Subtests
I	Allgemeine intellektuelle Fähigkeiten	3, 4, 5, 7, 8, 9, 10, 11, 12
II	Übliche Tätigkeiten von Lehrern	1+2, 6, 13, 14, 15

Marks, Seeber und Strümper (1975) verglichen die Faktorenstruktur verschiedener Berufsgruppen, namentlich 222 Locher und Prüfer (LP, $M = 29;4$ Jahre, $SD = 8;1$ Jahre), 101 Oberschüler der 10. Klasse (OS, $M = 16;4$ Jahre, $SD = 0;7$ Jahre), 108 Ärzte (Ä, $M = 31;8$ Jahre, $SD = 4;5$ Jahre) und 155 Lehrlinge der Datenverarbeitung (LD, $M = 16;1$ Jahre, $SD = 0;6$ Jahre). Neben einer rotierten Lösung gaben sie die Ergebnisse einer Transformationsanalyse zur besseren Vergleichbarkeit der Gruppen an. Kriterien zur Bestimmung der Faktorenzahl waren eine Lösung von 4 bis 6 Faktoren und die Beschränkung auf mindestens 5 % Varianzklärung. Da sich die Ergebnisse der einzelnen Berufe zwar ähneln, es aber zu spezifischen in der Studie diskutierten Unterschieden kam, werden alle nach der Transformationsanalyse rotierten Lösungen präsentiert (vgl. Tabelle 2.6). Die Autoren interpretierten und benannten zwar allgemein drei Faktoren für alle Stichproben; diese lassen sich aber nicht in allen Stichproben finden, so dass auf deren Erwähnung verzichtet wird.

Tabelle 2.6: Interpretationen der Faktorenanalyse der Untersuchung von Marks, Seeber und Strümper (1975), Lösung der Transformationsanalyse

	LP ($n = 222$)	OS ($n = 101$)	Ä ($n = 108$)	LD ($n = 155$)
I	1, 2, 5, 6, 12	1, 2, 5, 12	1, 2, 12, 13	1, 2, 5, 12
II	8, 9, 10	3, 8, 9, 11	7, 8, 9, 11	6, 7, 9, 11
III	3, 4	4, 10	3, 4, 5, 10	3, 4, 8, 10
IV	11, 13, 14	6, 13	6	13, 14
V	7	7	14	

Bemerkung: LP – Locher und Prüfer, OS – Oberschüler der 10. Klasse, Ä – Ärzte, LD – Lehrlinge der Datenverarbeitung

Reinert, Baltes und Schmidt (1965) verglichen zur Überprüfung der Leistungsdifferenzierungshypothese zwei Altersgruppen einmal parallelisiert nach gleicher Testleistung (A_1 : $M = 10;6$ Jahre, $Sp = [9;7, 11;6]$, $n = 137$; A_2 : $M = 12;6$ Jahre, $Sp = [11;7, 13;6]$, $n = 123$) und einmal parallelisiert nach gleichen Standardwerten (B_1 : $M = 10;1$ Jahre, $Sp = [9;7, 10;6]$, $n = 100$; B_2 : $M = 13;1$ Jahre, $Sp = [12;7, 13;6]$, $n = 100$). Sie führten allerdings nicht alle Subtests des LPS durch, da Subtest 2 genügend durch Subtest 1 und Subtest 4 durch 3 repräsentiert seien. Die

Angabe von Gründen für die Aussparung der Subtests 5, 7 und 8 fehlen. Neben dem LPS wurde das Begabungstestsystem (BTS; Horn, 1956) durchgeführt und in die Berechnung der Faktorenanalyse einbezogen. Die Ergebnisse werden hier ohne die Subtests des BTS präsentiert. Es ergaben sich für alle vier Stichproben nach dem Kriterium von Saunders (Cattell, 1952) drei Faktoren, die mittels einer Rotation zur orthogonalen Simple Structure nach Horst (1941) rotiert wurden (Tabelle 2.7).

Tabelle 2.7: Interpretationen der Faktorenanalyse der Untersuchung von Reinert, Baltes und Schmidt (1965), Rotation zur orthogonalen Simple Structure

	A₁ (n = 137)	A₂ (n = 127)	B₁ (n = 100)	B₂ (n = 100)
I	1, 6, 12	1, 6, 12	1, 6, 11, 12	1, 6, 12
II	3, 9, 10, 11	3, 9, 10, 11	9, 10	3, 9, 10, 11
III	13, 14, 15	13, 14, 15	3, 13, 14, 15	13, 14, 15

Tent (1969) hat in einer großen Studie zur Vorhersagbarkeit von Schulleistung auch das LPS angewandt und die Daten von 293 dreizehnjährigen Mädchen und Jungen ($M = 13;3$ Jahre, $SD = n. a.$) verrechnet. Die Hauptachsenmethode mit Varimaxrotation ergab sechs Faktoren. Als Kriterium wurde eine möglichst gute psychologische Interpretation der Rotation vorgegeben.

Tabelle 2.8: Interpretation der Faktorenanalyse der Untersuchung von Tent (1969, $n = 293$), Varimax-rotierte Lösung

	Interpretation	Subtests
I	(Allgemeiner) sprachlicher Faktor	1, 2, 5, 6, 12
II	Schlussfolgerndes Denken	3, 4, 13
III	Leistungsmotiviertheit	14, (15) ¹
IV	Schulisches Lernen	(6,) 9, 15
V	Wiedererkennen von Symbolen	7, (5, 12)
VI	Räumliches Vorstellen	8, (9,) 10, 11

Bemerkung: ¹Subtests in Klammern werden vom Autor zur Interpretation des Faktors herangezogen, laden aber auf einem anderen Faktor höher.

Weiss, Kemmler, Deisenhammer, Fleischhacker und Delazer (2003) benutzten zwar nur die Subtests 7, 8, 9 und 10 neben einigen anderen Tests, ihre Untersuchung ergab aber, dass alle Subtests gemeinsam mit dem Dreidimensionalen Würfeltest (Gittler, 1990) und dem Mentalen Rotationstest (Shepard & Metzler, 1971) auf einem Faktor laden, den sie „visual-spacial factor“ nannten. Die Stichprobe bestand aus 97 Psychologie- und Medizinstudenten mit einem Durchschnittsalter von 25;1 Jahren ($SD = 3;5$ Jahre).

Wewetzer (1968) untersuchte die Faktorenstruktur des LPS an zwei Schülergruppen. Zum einen an 100 Oberschülern und zum anderen an 100 Gewerbeschülern für Feinmechanik und Elektrotechnik mit einem Durchschnittsalter von 17-18 Jahren. Ferner teilte er die gesamte Stichprobe in zwei zufällige Gruppen. Die Hauptachsenanalyse mit Varimaxrotation ergab bei einer Extraktion bis 99 % der Varianz sechs Faktoren bei den Gewerbeschülern (GS), fünf bei den Oberschülern (OS) und je drei bei den beiden zufälligen Gruppen (ZuF₁, ZuF₂), wobei dann nicht alle Subtests essentielle Ladungen haben (Tabelle 2.9). Es wird allerdings nicht ersichtlich, welches Kriterium wirklich die Zahl der Faktoren bestimmt hat.

Tabelle 2.9: Interpretationen der Faktorenanalyse der Untersuchung von Wewetzer (1968), Varimax-rotierte Lösung

	GS (n = 100)	OS (n = 100)	ZuF₁ (n = 100)	ZuF₂ (n = 100)
I	1, 2, 5, 12	1, 2	1, 2, 5, 6, 12	1, 2, 5, 6, 12
II	4, 13	3, 4	3, 4, 13	3, 4, 10
III	8, 9, 10, 11	7, 8, 9, 11, 12	8, 9, 10, 11	8, 9, 10, 11
IV	6	5, 6, 10	(7, 14) ¹	(7, 13, 14)
V	14	13, 14		
VI	3, 7			

Bemerkung: ¹Subtests in Klammer laden auf keinen der drei Faktoren essentiell. Ob eine Vier-Faktorenlösung zu einem anderen Ergebnis geführt hätte, kann nicht nachvollzogen werden.

Zusammenfassend zeigt sich, dass gewisse Abhängigkeiten der Faktorenstruktur von Alter, sozio-ökonomischem Status und sicherlich weiteren Variablen, wie von Marks et al. (1975) konkret untersucht, existieren. Es zeigen sich aber auch gewisse Muster, die in vielen Untersuchungen zumindest Ähnlichkeiten aufweisen. In Tabelle 2.10 und Tabelle 2.11 werden die Studien noch einmal zusammengefasst.

Tabelle 2.10: Anzahl der Stichproben mit bestimmter Faktorenzahl

Faktoren	Stichproben	in Untersuchungen
2	1	1
3	7	3
4	2	2
5	6	3
6	2	2
7	1	1

Es wird deutlich, dass es die eine den LPS Subtests zugrunde liegende Struktur nicht geben kann. Zeigt Tabelle 2.10 bereits die Heterogenität hinsichtlich der Anzahl der Faktoren über verschiedene Stichproben hinweg auf, wird in Tabelle 2.11 deutlich, dass nur ein Teil der von Horn postulierten Subtests essentielle Gemeinsamkeiten aufweist. Dafür laden andere Subtests des LPS augenscheinlich häufig auf gemeinsamen Faktoren. Da der Vergleich aber nicht nur auf verschiedensten Stichproben beruht, sondern auch auf unterschiedlichen faktorenanalytischen Verfahren, unterschiedlichen Kriterien zur Bestimmung der Faktorenzahl und teilweise unterschiedlichen Rotationen der Faktorenmatrizen, werden im Folgenden Ergebnisse von Untersuchungen reanalysiert, die den LPS zum größten Teil durchführten und deren zugrunde liegende Interkorrelationsmatrizen zugänglich waren.

Tabelle 2.11: Zusammenfassung der Ergebnisse der vorgestellten Studien: Verhältnisse des gemeinsamen Ladens einzelner Subtests auf einem Faktor, rechts neben der Diagonalen als absolute Häufigkeiten, links als Prozentwerte

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		10/10	0/17	0/13	9/13	9/17	0/13	0/13	0/17	0/17	0/17	14/17	2/17	1/17	1/8
2	1		0/10	0/10	7/10	4/10	0/10	0/10	0/10	0/10	0/10	8/10	1/10	0/10	0/1
3	0	0		11/13	2/13	0/17	3/13	5/13	7/17	10/17	5/17	1/17	3/17	1/17	1/8
4	0	0	.85		2/13	0/13	2/13	4/13	3/13	7/13	1/13	1/13	3/13	0/13	0/4
5	.69	.70	.15	.15		6/13	1/13	1/13	1/13	3/13	1/13	10/13	1/13	0/13	0/4
6	.53	.40	0	0	.46		1/17	0/13	1/17	1/17	2/17	8/17	3/17	1/17	1/8
7	0	0	.23	.15	.08	.06		4/13	5/13	2/13	5/13	3/13	1/13	3/13	0/4
8	0	0	.38	.31	.08	0	.31		10/13	9/13	8/13	2/13	0/13	0/13	0/4
9	0	0	.41	.23	.08	.06	.38	.77		11/17	11/17	2/17	0/17	0/17	1/8
10	0	0	.59	.54	.23	.06	.15	.69	.65		8/17	1/17	0/17	0/17	0/8
11	0	0	.29	.08	.08	.12	.38	.62	.65	.47		4/17	1/17	2/17	0/8
12	.82	.80	.06	.08	.77	.47	.23	.15	.12	.06	.24		1/17	1/17	0/8
13	.12	.10	.18	.23	.08	.18	.08	0	0	0	.06	.06		10/17	6/8
14	.06	0	.06	0	0	.06	.23	0	0	0	.12	.06	.59		7/8
15	.13	0	.13	0	0	.13	0	0	.13	0	0	0	.75	.88	

Bemerkung: Grau hinterlegte Felder kennzeichnen Subtestpaarungen, die häufiger als 40 % auf demselben Faktor laden

Bergemann, Daumenlang und Johann (1981) reanalysierten Teile der Daten der Erhebung von Reinert et al. (1965) mittels einer Clusteranalyse und fanden einen Hinweis auf die Gültigkeit der Intelligenzdifferenzierungshypothese. Das clusteranalytische Ergebnis bei 10-jährigen konvergiert bei zwei, bei 13-jährigen bei drei Clustern (Complete-Linkage-Verfahren) bzw. drei und vier Clustern (Ward-Methode). Die Zusammensetzung der LPS-Subtests in den Clustern wird allerdings durch die BTS-Subtests beeinflusst, so dass hieraus nur bedingt Rückschlüsse auf die alleinige Struktur des LPS gezogen werden können. Die gewählten Partitionen sind in Tabelle 2.12 zusammengestellt. Hierbei wird deutlich, dass die so gefundene Struktur der Subtests nur in einem sehr begrenzten Rahmen der der Faktorenanalysen entspricht.

Tabelle 2.12: Clusteranalytische Ergebnisse der Reanalyse der Daten von Reinert et al. (1965) durch Bergemann et al. (1981)

	CL 10-jährige	CL 13-jährige	Ward 10-jährige	Ward 13-jährige
I	1, 3, 6, 9, 10, 12, 13, 15	1, 3, 9	1, 3	1, 3
II	11, 14	6, 10	6, 10, 12, 15	6, 9, 10
III		11, 12, 13, 14, 15	9, 11, 13, 14	12, 15
IV				11, 13, 14

Bemerkung: Zusätzlich zu den LPS-Subtests werden in der Originalliteratur die Subtests des BTS angegeben.

Steck et al. (1975) berechneten drei Faktorenanalysen (Hauptachsenanalyse mit Rotation zur Einfachstruktur nach Überla, 1971) mit dem LPS und weiteren Intelligenztests, um die Gültigkeit der postulierten Intelligenzstrukturen zu prüfen: 1. LPS, IST, HAWIE mit 111 Schülern und Studenten, Lösung mit fünf Faktoren; 2. LPS, CFT, HAWIE, d2 mit den 111 Probanden plus 31 freiwilligen Bundeswehrangehörigen, Lösung mit vier Faktoren; 3. LPS, IST, EVT (Eignungs- und Verwendungstest der Bundeswehr) mit 187 Fachoffiziersanwärtern, Lösung mit sechs (LPS-Subtests laden nur auf vier) Faktoren. Die Ergebnisse sind in Tabelle 2.13 zusammengefasst, wobei auch hier mit einer Wechselwirkung mit den anderen verrechneten Testergebnissen auszugehen ist.

Tabelle 2.13: Ergebnisse der Faktorenanalysen von Steck et al. (1975)

	1. (n = 111)	2. (n = 142)	3. (n = 187)
I	1+2, 5, 6	1+2, 5, 6	1, 2, 5
II	3, 4, 9, 10	3, 7, 8, 9, 10, 11	3, 4, 7
III	7, 13, 14	4	6
IV	8, 11	12, 13, 14	8, 9, 10
V	12		

Bemerkung: zu 1. Faktorenanalyse zusätzlich mit IST und HAWIE
zu 2. Faktorenanalyse zusätzlich mit CFT, HAWIE und d2
zu 3. nur LPS-Subtests 1-10; Faktorenanalyse zusätzlich mit IST und EVT resultiert in sechs Faktoren

Steck et al. (1975) interpretierten ihre Ergebnisse dahingehend, dass sich unter Berücksichtigung aller Tests die Intelligenzdimensionen „Sprachbeherrschung“ (LPS: 1+2, 5, 6) und „anschauungsgebundenes Denken“ (LPS: 8, 9 und etwas schwächer 7, 10, 11) sowie „Rechenfähigkeit“ (kein LPS-Subtest), „Konzentrationsvermögen“ (LPS: 13, 14 + 7, 12) und „Schlussfolgerndes Denken“ (LPS: 3, 4) finden lassen.

2.2 Reanalyse einzelner Untersuchungen

In einigen der Untersuchungen mit einer genügend großen Anzahl von Subtests des LPS sind die Interkorrelationsmatrizen der Subtests publiziert, so dass eine erneute Berechnung einer Faktorenanalyse unter gleichen methodischen Bedingungen möglich ist. Im Folgenden werden zuerst drei von Horn im Manual des LPS (1983) veröffentlichte Datensätze neu ausgewertet, anschließend die zum Teil bereits oben erwähnten Daten und teilweise Daten aus Studien, in denen selbst keine Analyse der faktoriellen Struktur vorgenommen wurde. Die Ergebnisse basieren auf einer Hauptkomponentenanalyse. Diese wurde gewählt, da dabei a) versucht wird, einen möglichst großen Teil der Varianz der Daten zu klären, und b) es das meist benutzte und bekannteste Verfahren ist (Gaensslen & Schubö, 1973). Als Kriterium für die Anzahl der extrahierten Faktoren wurde das Kaiser-Guttman-Kriterium (Eigenwert > 1) gewählt, da dadurch sichergestellt wird, dass die durch einen Faktor erklärte Varianz größer ist, als es allein durch eine einzelne Variable möglich wäre (Backhaus, Erichson, Plinke & Weiber, 2003). Nach der Extraktion werden die Ergebnisse mittels Varimax-Methode (Kaiser, 1958) rotiert, um eine möglichst gute Anpassung an eine Einfachstruktur bei Erhalt der Orthogonalität zu erreichen (Gaensslen & Schubö, 1973). Da hierbei keine genaue Analyse der Faktorenanalyse im Vordergrund steht, sondern v. a. die Faktorenstruktur des LPS in unterschiedlichen Stichproben, werden als Ergebnisse wie bereits in Kapitel 2.1 nur die Faktoren und die auf sie am höchsten ladenden Subtests nach der Rotation aufgeführt.

Der Interkorrelationstabelle von Horn (1983) liegen die Daten von 200 erwachsenen Probanden zu Grunde. Koopmann (1964, zit. nach Horn, 1983) untersuchte 159 Volksschüler der 4. Klasse und 159 Pioniere der Bundeswehr (Tabelle 2.14).

Tabelle 2.14: Interpretationen der Faktorenanalyse der Daten aus dem Manual zum LPS (Horn, 1983; Koopmann, 1964)

	Horn (n = 200)	Volksschüler (n = 159)	Pioniere der Bundeswehr (n = 159)
I	1+2, 4, 5, 6, 12, 13, 14, 15	1+2, 5, 6, 12	1+2, 4, 5, 6, 12, 13, 14
II	3, 7, 8, 9, 10, 11	3, 4	3, 7, 8, 9, 10, 11
III		8, 9, 10	
IV		7, 11, 13, 14	

Die Interkorrelationstabellen der Untersuchung von Bittner (1960) stellen insofern eine Ausnahme dar, da sie auf einer Vorform des LPS basieren. Sie sollen zwar auch hier Erwähnung finden, ihre Ergebnisse müssen aber unter Vorbehalt betrachtet werden (Tabelle 2.15).

Tabelle 2.15: Interpretationen der Faktorenanalyse der Daten von Bittner (1960)

	labile (n = 259)	stabile (n = 262)
I	1, 2, 4, 5, 6, 12, 15	1, 2, 6, 12
II	3, 7, (8), 9, 10, 11	3, 4, 5, 10, 15
III		7, (8), 9, 11

Bemerkung: Die Nummerierung der Subtests wurde der aktuellen Version angepasst. Subtest 13 und 14 gab es in der Vorform noch nicht. Subtest 8 entsprach einer anderen Itemart.

Die Reanalyse der Daten von Greif (1972) ergibt vier und nicht wie bei Greif fünf Faktoren (Tabelle 2.16). Die Reanalyse der Daten von Kury (1972) führt hingegen zum gleichen Ergebnis wie die Originaluntersuchung. Die dritte in Tabelle 2.16 angeführte Untersuchung stammt von Schumacher (1988), der die Ergebnisse der ersten zehn Subtests des LPS von 160 Personen zusammenstellte. Die Daten stammen von 80 Arbeitslosen und 80 Arbeitenden, je 40 erhoben im Jahr 1974 und 1984. Im Originalartikel wurde keine Faktorenanalyse durchgeführt.

Tabelle 2.16: Interpretationen der Faktorenanalyse der Daten von Greif (1972), Kury (1972) und Schumacher (1988)

	Greif (n = 302)	Kury (n = 187)	Schumacher (n = 160)
I	1+2, 6, 11, 12	1, 2	1+2, 5, 6
II	3, 7, 8, 9, 10	3, 4, 8, 9, 10	3, 4, 7, 8, 9, 10
III	4, 5	5, 6, 13	
IV	13, 14, 15	7, 11, 12, 14	

Tent (1969) folgte aus der Faktorenanalyse der 293 Volksschüler fünf Faktoren. Die Reanalyse ergab hingegen nur vier. Ferner werden in der Veröffentlichung die Interkorrelationsmatrizen der LPS-Skalen 1-14 von 74 Gymnasiasten, 103 Realschülern, 347 Mädchen und 295 Jungen der Volksschule angegeben. Diese Werte wurden ebenfalls für die Berechnung von Faktorenanalysen herangezogen. Hierbei ergaben sich eine Lösung mit drei, zwei mit vier und eine mit fünf Faktoren (Tabelle 2.17).

Tabelle 2.17: Interpretationen der Faktorenanalyse der Daten von Tent (1969)

	Volksschüler (n = 293)	Mädchen Volksschule (n = 347)	Jungen Volksschule (n = 295)	Realschüler (n = 103)	Gymnasiasten (n = 74)
I	1, 2, 5, 6, 12	1, 2, 5, 6	1, 2, 5, 6, 11	1, 2, 5	1, 2, 5
II	3, 4, 9, 13	3, 4, 8, 9, 10	3, 4	3, 4	3, 4, 8, 13
III	7, 8, 10, 11	7, 11	7, 8, 9, 10, 12, 13, 14	6, 8, 10	6, 10, 11, 12, 14
IV	14, 15	12, 13, 14		7, 11, 12, 14	7, 9
V				9, 13	

Resultieren in der Untersuchung von Reinert et al. (1965) für jede Stichprobe drei Faktoren, finden sich bei der Reanalyse für drei der vier Gruppen (Ausnahme B₁) vier Faktoren (vgl. Tabelle 2.18). Ebenso resultieren vier Faktoren für die Gesamtgruppe aus A₁ und A₂ und drei Faktoren für B₁ und B₂, die sich durch Zusammenlegung der Korrelationsmatrizen ergab. Dies war möglich, da für jeden Subtest die Varianz in den einzelnen Gruppen angegeben war. Dadurch konnten die Abweichungsquadrate der Variablenpaare und daraus wiederum die Kovarianzmatrix als Grundlage der Faktorenanalyse berechnet werden.

Tabelle 2.18: Interpretationen der Faktorenanalyse der Daten von Reinert et al. (1965)

	A₁ (n = 137)	A₂ (n = 127)	B₁ (n = 100)	B₂ (n = 100)	A₁+A₂ (n = 260)	B₁+B₂ (n = 200)
I	1, 12	1, -14	1, 12	1, 12	1, 12, -14	1, 11, 12
II	3, 6, 10	3, -6, 9, 10	3, 6, 9, 10, 11	3, 10	-3, 6	3, 9, 10
III	9, 11	11, 12	13, 14, 15	9, 11, 14	9, 10, 11	6, 13, 14, 15
IV	13, 14, 15	13, 15		6, 13, 15	13, 15	

Bemerkung: „-“ vor dem Subtest bedeutet, dass der Test negativ mit dem Faktor korreliert.

Ein weiteres Ziel der Reanalyse der Daten wäre gewesen, durch die Zusammenlegung möglichst verschiedener Daten eine Faktorisierung der LPS-Skalen in einer sehr heterogenen Stichprobe vorzunehmen. Dies war aber aus einigen Gründen nicht möglich. Zum einen ließen sich nur zwei Studien finden, in denen alle 15 Subtests des LPS und die dazugehörige Interkorrelationsmatrix der Subtests veröffentlicht waren. In keiner der beiden wurden vollständig die Standardabweichungen der Subtests angegeben, so dass eine Weiterverarbeitung der Daten nicht möglich ist. Einige Studien fassen die Subtests 1 und 2 zusammen, andere nicht. In diesem Fall ist eine Zusammenfassung oder Trennung ohne die Rohdaten nicht möglich. In einigen Arbeiten wurde nur ein Teil der Subtests veröffentlicht bzw. benutzt, so dass die Schnittmenge im Resultat sehr klein und daher wenig aussagekräftig wäre.

Um ein noch besseres Bild von den Zusammenhängen der Subtests zu erhalten, wurden Studien gesucht, in welchen die Daten des alten PSB (Prüfsystem für Schul- und Bildungsberatung, Horn, 1969) veröffentlicht sind, da bei diesem Test neun der zehn Skalen den gleichen Itemtyp enthalten wie im LPS. Der Einfachheit halber werden im Folgenden nicht die Bezeichnungen des PSB sondern die der entsprechenden LPS-Skalen benutzt (vgl. Tabelle 2.19).

Tabelle 2.19: Sich entsprechende Subtests von PSB und LPS

PSB	1	2	3	4	5	6	7	8	9	10
LPS	1	2	3	4	6	12	8	10	-	14

Es wurden drei Studien gefunden, die eine Reanalyse der Daten durch die Veröffentlichung der Interkorrelationsmatrizen ermöglichten: Aurin (1968) testete in einer Studie zu Bildungschancen in Baden-Württemberg u. a. 473 Schüler der 4. Klassen einer Region mit geringer Bildungsdichte (Tabelle 2.20).

Tabelle 2.20: Interpretationen der Faktorenanalyse der Daten von Aurin (1968)

	Schüler der 4. Klasse ($n = 473$)
I	1+2, 6, 12, (9)
II	3, 4, 8, 10, 14

Steller und Stürmer (1984) untersuchten drei Stichproben mit dem PSB, um den Nachweis zu erbringen, dass der LPS-Subtest 8 (PSB 7) nicht dem Faktor „Space“, sondern der Primary Mental Ability „Reasoning“ zuzuordnen sei. Die erste Stichprobe (EB) bilden 49 Kinder ($M = 11;11$, $SD = 1;3$), die in den Erziehungsberatungsstellen der Stadt Kiel getestet wurden. Die zweite Stichprobe (OR) bestand aus 41 Kindern ($M = 12;3$, $SD = 0;9$), die im Rahmen einer Schulversetzung aus dem Gymnasium bzw. aus der Realschule in den jeweils niedrigeren Schultyp getestet wurden. Als dritte Stichprobe (DEL) wurden die Daten von 74 Jugendlichen ($M = 19;7$, $SD = 1;5$) gewählt, die im Rahmen der Entlassungsvorbereitungen aus der Jugendabteilung der Justizvollzugsanstalt Neumünster getestet wurden. Aufgrund der Veröffentlichung der Standardabweichungen der einzelnen Stichproben in den Subtests konnten die drei Gruppen zudem zu einer zusammengefasst und analysiert werden (Tabelle 2.21).

Tabelle 2.21: Interpretationen der Faktorenanalyse der Daten von Steller und Stürmer (1984)

	EB ($n = 74$)	OR ($n = 49$)	DEL ($n = 74$)	Alle ($n = 164$)
I	1+2, 6, 12	1+2, 6, 12	1+2, 4, 6, (9), 12	1+2, 6, 12
II	3, 4, 8, 10	3, 4, 8, 10	3, 8, 10, 14	3, 4, 8, 10
III	(9), 14	(9), 14		(9), 14

Bemerkung: EB – Getestet in der Erziehungsberatungsstelle, OR – Getestet im Rahmen eines Schulwechsels, DEL – Getestet zur Entlassungsvorbereitung aus der Jugend-JVA

Langfeldt-Nagel (1979) untersuchte den Einfluss der Schulbildung auf Intelligenztestleistungen und testete zu diesem Zweck 185 Hauptschüler und 294 Gymnasiasten. Diese wurden zum einen bereits in der 4. Klasse getestet, ein zweites Mal sieben Jahre später. Aus den Angaben der Arbeit konnten acht Interkorrelationsmatrizen und zwei Zusammenfassungen gewonnen werden, wobei diese Ergebnisse aufgrund der wiederholten Testung a priori nicht unabhängig sein können (Tabelle 2.22).

Tabelle 2.22: Interpretationen der Faktorenanalyse der Daten von Langfeldt-Nagel (1979)

	HS-1-♀ (n = 48)	HS-1-♂ (n = 85)	Gym-1-♀ (n = 123)	Gym-1-♂ (n = 171)	Gesamt-1 (n = 427)
I	1+2, 6, 12, (9)	1+2, 8, 10, 12	1+2, 12	1+2, 6, 12	1+2, 6, 12
II	3, 4, 10	3, 4	3, 4, 8, 10	3, 4, 8, 10	3, 4, 8, 10
III	8, 14	6	6, (9), 14	(9), 14	(9), 14
IV		(9), 14			
	HS-2-♀ (n = 48)	HS-2-♂ (n = 85)	Gym-2-♀ (n = 123)	Gym-2-♂ (n = 171)	Gesamt-2 (n = 427)
I	1+2, 6, 12	1+2, 6, 12	1+2, 6, (9), 14	1+2, 6, 12	1+2, 6, (9), 14
II	3, 4, 8, 10	3, 4, (9), 14	3, 4, 8, 10	3, 4, 8, 10	3, 4, 8, 10
III	(9), 14	8, 10		(9), 14	
IV					

Bemerkung: HS – Hauptschule, Gym – Gymnasium, 1 – Eingangsuntersuchung 4. Klasse, 2 – Enduntersuchung nach sieben Jahren, ♀ – Mädchen, ♂ – Jungen

Die deutlich werdende faktorielle Struktur des PSB und damit der Zusammenhang der LPS-Skalen entsprechen nicht ganz der von Horn (1969) postulierten Struktur. Es zeigt sich, dass die verbalen Subtests 1, 2, 6 und 12 sehr deutlich auf einem Faktor laden. Die Subtests 3 und 4 sowie 8 und 10 laden oft auf einem Faktor, so dass eine inhaltliche Trennung der Konstrukte als nicht angemessen erscheint. Die Subtests 9 und 10 scheinen einen dritten Faktor darzustellen, auch wenn einige Male keine ausreichende Trennung der Fähigkeiten möglich war und nur zwei Faktoren resultierten.

Tabelle 2.23: Zusammenfassung der Ergebnisse der reanalysierten Daten: Verhältnisse des gemeinsamen Ladens einzelner Subtests auf einem Faktor, rechts neben der Diagonalen als absolute Häufigkeiten, links als Prozentwerte

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1		8/8	0/19	3/13	10/13	10/19	0/13	0/11	0/19	0/19	1/18	12/18	2/16	4/16	1/11
2	1*		0/7	1/9	6/8	5/8	0/8	0/8	0/8	0/8	1/8	3/8	0/6	0/6	0/3
3	0	0		9/13	1/13	4/19	5/13	6/11	11/19	13/19	4/18	0/18	2/16	0/16	0/11
4	.23	.11	.69*		5/13	3/13	1/13	4/11	4/13	4/13	0/12	2/12	1/10	0/10	3/5
5	.77*	.75*	.08	.38		9/13	0/13	0/11	0/13	1/13	1/12	5/12	3/10	2/10	3/5
6	.53*	.63*	.21	.23	.69*		0/13	0/11	2/19	5/19	3/18	8/18	5/16	4/16	4/9
7	0	0	.38	.08	0	0		6/12	8/13	7/13	9/12	3/12	2/10	4/10	0/5
8	0	0	.54**	.36	0	0	.50**		8/11	10/11	3/10	0/10	2/10	1/10	0/3
9	0	0	.58*	.31	0	.11	.62**	.73*		13/19	8/18	1/18	3/16	2/16	0/11
10	0	0	.68*	.31***	.08	.26	.54**	.91*	.68*		7/18	2/18	1/16	2/16	0/11
11	.06	.13	.22	0	.08	.17	.75**	.30***	.44*	.38***		6/18	1/16	5/16	0/11
12	.66*	.38***	0	.17	.42*	.44*	.25	0	.06	.13	.33		4/16	8/16	2/11
13	.13	0	.13	.10	.30	.31	.20	.20	.19	.06	.06	.25		9/16	8/9
14	.25	0	0	0	.20	.25	.40**	.10	.13	.13	.31	.50	.56*		6/9
15	.09	0	0	.60**	.60**	.44**	0	.0	0	0	0	.20	.89*	.66*	

Bemerkung: * Subtestpaarungen, die sowohl in den Literaturangaben als auch bei der Reanalyse häufiger als 40 % vorkommen.

** Subtestpaarungen, die nur in der Reanalyse häufiger als 40 % vorkommen.

*** Subtestpaarungen, die nur in den Literaturangaben häufiger als 40 % vorkommen.

Beim Vergleich der Ergebnisse der Reanalyse mit der Zusammenfassung der Ergebnisse aus der Literatur wird deutlich, dass ein relativ strenges Kriterium für die Anzahl der Faktoren wie das Kaiser-Guttman-Kriterium nur in einem Fall zu einem Ergebnis mit mehr als vier Faktoren

führt (vgl. Tabelle 2.24). Dadurch bedingt ist eine höhere Zahl von Subtestpaarungen, die relativ häufig auf demselben Faktor laden (vgl. Tabelle 2.23).

Tabelle 2.24: Anzahl der Stichproben mit bestimmter Faktorenzahl in der gefundenen Literatur und der Reanalyse

Faktoren	Stichproben (Literatur)	in unterschiedlichen Untersuchungen	Stichproben (Reanalyse)	in unterschiedlichen Untersuchungen
2	1	1	4	4
3	7	3	4	3
4	2	2	10	5
5	6	3	1	1
6	2	2	-	-
7	1	1	-	-

Im Folgenden werden die gefundenen Korrelationen der Subtests mit einzelnen externen Kriterien sowie anderen Testergebnissen behandelt, bevor abschließend die Bewertung der Ergebnisse und die gezogenen Schlussfolgerungen diskutiert werden.

2.3 Korrelative Zusammenhänge

Zur Beurteilung der Skalen eines Tests, auch im Hinblick auf eine Revision und Aussortierung einzelner Subtests als unnötig, sollten die Zusammenhänge der Subtests mit anderen Tests und mit externen Kriterien Beachtung finden. Im Falle des LPS ließen sich Untersuchungen finden, die Korrelationen des LPS mit den Subskalen des IST (Amthauer, 1955; 1970), des WIT (Jäger, 1963) und des AIT (Meili, 1966), und einzelne mit KLT (Düker & Lienert, 1965), d2 (Brickenkamp, 1972) und FRT (Daniels, 1971) enthielten. Als externe Kriterien wurden ausschließlich schulische Leistungsmaße gefunden.

2.3.1 Korrelationen mit anderen Tests

Groffmann und Schneevoigt (1964) testeten 31 männliche und 6 weibliche Studierende einer Wirtschaftshochschule ($M = 24;2$ Jahre, $Sp = [21, 32]$) sowohl mit dem LPS (Horn, 1962) als auch mit dem IST (Amthauer, 1955). In Tabelle 2.25 wird die Interkorrelationsmatrix der Untertests präsentiert.

Tabelle 2.25: Interkorrelationskoeffizienten zwischen Untertests des LPS und IST ($n = 37$, aus Groffmann & Schneevoigt, 1964, S. 249)

	1+2	3	4	3+4	5	6	5+6	7	8	9	10	7-10	11	12	11+12	13	14	13+14
SE	-.07	.35	.28	.31	.15	-.19	.08	.19	.15	.25	.25	.58	.19	.09	.29	.14	.00	-.10
WA	.19	.45	.96	.52	.18	.07	.17	.38	.53	.47	.38	.60	.38	.44	.58	.17	.22	.07
AN	.34	.56	.45	.45	.38	.40	.35	.42	.56	.42	.42	.60	.37	.83	.48	.22	.07	.06
GE	.20	.29	.27	.28	.41	-.04	.35	.44	.21	.06	.39	.60	.12	.85	.36	.06	-.02	-.14
ME	.31	.39	.36	.28	.28	.14	.27	.23	.44	.41	-.14	.52	.41	.49	.49	.12	.12	.01
RA	.42	.67	.45	.61	.41	.00	.32	.22	.41	.44	.11	.59	.09	.44	.36	.00	.11	-.10
ZR	.18	.59	.43	.56	.40	.12	.35	.14	.32	.35	-.47	.66	.20	.38	.40	.05	.24	.01
FA	-.95	.50	.42	.51	.05	.16	.05	.15	.50	.32	-.40	.64	-.04	.39	.30	-.17	.14	-.25
WÜ	-.12	.51	-.72	.60	.13	.30	.19	.36	.52	.11	-.70	.76	.30	.16	.35	.25	-.07	-.03

Bemerkung: $r > .32$ sign. 5%, $r > .42$ sign. 1%, SE – Satzergänzungen, WA – Wortauswahl, AN – Analogien, GE – Gemeinsamkeiten, ME – Merkaufgaben, RA – Rechenaufgaben, ZR – Zahlenreihen, FA – Figurenauswahl, WÜ – Würfelaufgaben

Die Korrelationstabelle der Untersuchung von Kury (1972, Tabelle 2.26) zwischen LPS und IST (Amthauer, 1955) weist eine prinzipiell ähnliche Struktur auf wie die oben genannte von Groffmann und Schneevoigt (1964), wenngleich in der ersten auffällig mehr negative Koeffizienten präsentiert werden.

Tabelle 2.26: Interkorrelationskoeffizienten zwischen Untertests des LPS und IST ($n = 187$, aus Kury, 1972, S. 64)

	1	2	1+2	3	4	3+4	5	6	5+6	7	8	9	10	7-10	11	12	11+12	13	14	13+14
SE	.26	.32	.30	.19	.19	.23	.22	.24	.29	.23	.09	.21	.24	.26	.13	.14	.17	-.01	-.07	-.06
WA	.26	.21	.25	.21	.19	.24	.25	.08	.20	.25	.13	.22	.17	.26	.31	.11	.27	.05	.12	.12
AN	.35	.39	.39	.37	.26	.35	.26	.23	.31	.32	.25	.24	.32	.38	.22	.27	.28	.03	-.05	-.02
GE	.32	.33	.34	.28	.20	.28	.22	.18	.26	.18	.13	.20	.18	.22	.32	.30	.37	-.02	.13	.05
ME	.21	.23	.24	.28	.30	.31	.23	.25	.29	.17	.29	.34	.32	.37	.18	.13	.20	.23	.17	.26
RA	.32	.31	.32	.49	.45	.54	.39	.26	.39	.24	.33	.44	.30	.43	.12	.23	.20	.02	.12	.10
ZR	.26	.23	.26	.43	.48	.51	.41	.31	.42	.26	.20	.40	.34	.38	.15	.19	.19	.08	.34	.27
FA	.33	.34	.34	.42	.37	.45	.27	.24	.30	.24	.45	.44	.35	.50	.28	.31	.36	.03	.04	.05
WÜ	.09	.08	.09	.40	.32	.41	.20	.09	.17	.18	.44	.46	.25	.44	.20	.24	.26	.04	.15	.13

Bemerkung: $r > .14$ sign. 5 %, $r > .19$ sign. 1 %, SE – Satzergänzungen, WA – Wortauswahl, AN – Analogien, GE – Gemeinsamkeiten, ME – Merkaufgaben, RA – Rechenaufgaben, ZR – Zahlenreihen, FA – Figurenauswahl, WÜ – Würfelaufgaben

Greif (1972) korreliert in der Untersuchung der 302 Bereitschaftspolizisten die LPS-Skalen mit den Subtests der Intelligenztests WIT (Wilde-Intelligenz-Test, Jäger, 1963), IST (Amthauer, 1955) und AIT (Analytischer Intelligenztest, Meili, 1966) (Tabelle 2.27a - c). In allen Untersuchungen wird deutlich, dass die Subtests 5, 6, 7, 11, 12, 13 und 14 nur in begrenztem Maße mit den Untertests anderer Intelligenztests korrelieren. Die Reasoning-Subtests 3 und 4, die Space-Subtests 8 und 9 sowie Subtest 10, der auch räumliche Wahrnehmung i. w. S. fordert, scheinen hingegen entsprechende Pendanten in den anderen Intelligenztests zu haben. Die Subtests 1+2 korrelieren nur vereinzelt bedeutsam mit sprachrelevanten Untertests. Die Korrelation mit dem Subtest Figurenauswahl des IST von -.95 erscheint eher als Druckfehler, da dieses Ergebnis bei Groffmann und Schneevoigt (1964) keine Erwähnung findet und bei Greif (1972) nicht repliziert wird.

Tabelle 2.27a: Interkorrelationskoeffizienten zwischen den Untertests des LPS und den Untertests des IST ($n = 302$, aus Greif, 1972, S. 360)

	1+2	3	4	5	6	7	8	9	10	11	12	13	14
SE	.29	.19	.17	.20	.05	.15	.15	.20	.17	.09	.20	.07	.07
WA	.23	.27	.23	.21	.13	.14	.22	.26	.27	.11	.14	.07	.12
AN	.31	.39	.30	.24	.10	.15	.28	.26	.35	.15	.26	.03	.05
GE	.40	.27	.21	.25	.23	.20	.27	.16	.29	.30	.21	.07	.10
RA	.31	.43	.43	.34	.06	.21	.25	.33	.31	.09	.24	.19	.24
ZR	.24	.52	.52	.40	.14	.28	.34	.42	.38	.07	.34	.17	.33
FA	.15	.36	.25	.17	.08	.25	.44	.40	.35	.24	.26	.11	.17
WÜ	.13	.32	.28	.22	.09	.22	.41	.33	.29	.22	.21	.09	.12

Bemerkung: $r > .11$ sign. 5 %, $r > .15$ sign. 1 %, SE – Satzergänzungen, WA – Wortauswahl, AN – Analogien, GE – Gemeinsamkeiten, RA – Rechenaufgaben, ZR – Zahlenreihen, FA – Figurenauswahl, WÜ – Würfelaufgaben

Tabelle 2.27b: Interkorrelationskoeffizienten zwischen den Untertests des LPS und den Untertests des WIT ($n = 302$, aus Greif, 1972, S. 364)

	1+2	3	4	5	6	7	8	9	10	11	12	13	14
GR	.16	.16	.31	.24	.12	.14	.12	.11	.14	.08	.19	.19	.23
GW	.51	.18	.14	.29	.25	.19	.12	.15	.23	.19	.28	.06	.16
AL	.36	.37	.31	.21	.13	.15	.33	.35	.34	.15	.36	.05	.03
SCH	.29	.42	.45	.30	.13	.24	.21	.27	.31	.13	.22	.25	.36
SW	.33	.31	.24	.30	.25	.19	.26	.28	.23	.17	.25	.05	.21
ZN	.31	.37	.40	.31	.09	.17	.25	.27	.28	.06	.27	.20	.30
SP	.19	.43	.35	.17	.08	.39	.33	.41	.33	.23	.25	.14	.33
ER	.26	.32	.34	.29	.12	.20	.22	.26	.23	.10	.19	.19	.24
BR	.27	.48	.52	.29	.12	.25	.38	.34	.34	.22	.28	.19	.22
BO	.17	.43	.36	.25	.06	.24	.33	.28	.38	.32	.22	.28	.33
AW	.19	.37	.33	.17	.10	.18	.41	.37	.43	.31	.22	.01	.17
WG	.43	.25	.27	.27	.51	.25	.16	.21	.22	.28	.35	.13	.18

Bemerkung: $r > .11$ sign. 5 %, $r > .15$ sign. 1 %, GR – Grundrechnen, GW – Gleiche Wortbedeutung, AL – Analogien, SCH – Schätzen, SW – Sprichwörter, ZN – Zahlenreihen, SP – Spiegelbilder, ER – Eingekleidete Rechenaufgaben, BR – Buchstabenreihen, BO – Beobachtungen, AW – Abwicklungen, WG – Wortgewandtheit

Tabelle 2.27c: Interkorrelationskoeffizienten zwischen den Untertests des LPS und den Untertests des AIT ($n = 302$, aus Greif, 1972, S. 356)

	1+2	3	4	5	6	7	8	9	10	11	12	13	14
BR	.16	.46	.33	.21	.11	.25	.40	.38	.36	.29	.19	.15	.22
ZR	.25	.46	.49	.30	.19	.29	.35	.35	.37	.11	.32	.14	.30
SÄ	.31	.23	.21	.18	.41	.14	.20	.22	.24	.25	.22	.03	.19
LÜ	.26	.34	.26	.21	.15	.22	.33	.26	.30	.40	.27	.06	.19
ZE	.17	.35	.19	.15	.23	.16	.26	.27	.27	.17	.13	.14	.18
AL	.19	.40	.37	.17	.06	.22	.38	.37	.39	.19	.26	.13	.11

Bemerkung: $r > .11$ sign. 5 %, $r > .15$ sign. 1 %, BR – Bilderreihen, ZR – Zahlenreihen, SÄ – Sätze, LÜ – Lücken, ZE – Zeichnen, AL – Analogien

Schumacher (1988) verglich die Testergebnisse von LPS (Subtests 1-10), KLT, FRT und d2 von je 80 Arbeitslosen und „Arbeitshabenden“, die durch die Bundesanstalt für Arbeit 1974 und 1984 erhoben worden waren. Aus den angegebenen Daten lässt sich die Korrelationsmatrix der Tests bestimmen (Tabelle 2.28).

Tabelle 2.28: Interkorrelationskoeffizienten zwischen den Untertests des LPS mit dem KLT, dem FRT und dem d2 ($n = 160$, nach Schumacher, 1988)

	1+2	3	4	5	6	7	8	9	10
KLT_L	.35	.28	.36	.46	.37	.28	.30	.10	.27
KLT_F	-.14	-.15	-.17	-.19	-.01	-.16	-.21	-.11	-.11
FRT_L	.35	.62	.67	.34	.32	.42	.60	.23	.56
FRT_F	-.18	-.37	-.50	-.22	-.21	-.15	-.38	-.13	-.39
d2_L	.38	.31	.31	.45	.43	.33	.28	.08	.38
d2_F	-.04	-.25	-.27	-.14	.02	-.15	-.18	-.10	-.14

Bemerkung: $r > .16$ sign. 5 %, $r > .20$ sign. 1 %, _L bezeichnet die Anzahl bearbeiteter Items, _F die Zahl der Fehler

Die Leistungsmaße der Tests korrelieren mit Ausnahme von Subtest 9 mit KLT und d2 alle positiv und sehr hoch mit den Subtests des LPS. Für den FRT als Maß der allgemeinen Intelligenz waren ähnliche Ergebnisse zu erwarten, aber auch der Zusammenhang zwischen Aufmerksamkeit und der Bearbeitung der LPS-Subtests liegt augenscheinlich auf der Hand. Die Korrelation von Sorgfältigkeit im Sinne der Anzahl der Fehler mit den Subtests scheint auch als Ergebnis des Testaufbaus gewertet werden zu müssen.

2.3.2 Korrelationen mit schulischen Leistungsmaßen

Weiss (1972) veröffentlichte die Zusammenhänge der nach Horn zusammengefassten Subtests des LPS mit den Noten der „Hauptgegenstände“ Deutsch, Mathematik, Englisch und der 2. Fremdsprache sowie der schulischen Gesamtleistung von 200 Schülern und 82 Schülerinnen der 8. Klassen höherer Schulen (entspricht dem deutschen Gymnasium) (vgl. Tabelle 2.29). Es zeigte sich, dass der Gesamtscore mit allen außer den Fremdsprachennoten von Mädchen signifikant korreliert. Auffällig sind die geringe Zahl der bedeutsamen Korrelationen der Subtests 11+12, 13+14 sowie 7-10. Diese Ergebnisse weisen darauf hin, dass die Subtests nur geringe Validität bezüglich des Erfolgs in einzelnen Schulfächern zu haben scheinen.

Tabelle 2.29: Interkorrelationskoeffizienten zwischen den Untertests des LPS und einzelnen Schulleistungen in höheren Schulen ($n = 282$, nach Weiss, 1972)

	1+2	3+4	5+6	7-10	11+12	13+14	15	G
Schulische Gesamtleistung männl. Vpn	.20**	.17*	.08	.22**	.05	.10	.17*	.21**
Schulische Gesamtleistung weibl. Vpn	.17	.07	.11	.12	.21*	.15	.18	.21*
Deutsch männl. Vpn	.31**	.14*	.14*	.13	.08	.13	.16*	.21**
Deutsch weibl. Vpn	.38**	.24*	.37**	.17	.13	.14	.29**	.25*
Mathematik männl. Vpn	.10	.20**	.04	.17*	-.09	.13	.16*	.16*
Mathematik weibl. Vpn	.02	.18	.17	.20	.15	.20	.18	.29**
Englisch (1. Fremdsprache) männl. Vpn	.22**	.07	.18**	.13	.08	.03	.19**	.16*
Englisch (1. Fremdsprache) weibl. Vpn	.02	-.18	-.12	-.11	-.02	.16	.02	-.08
2. Fremdsprache ¹ männl. Vpn	.18**	.11	.05	.10	.05	.09	.10	.16*
2. Fremdsprache ¹ weibl. Vpn	-.04	-.28**	-.04	-.18	.00	.16	.02	-.19

Bemerkung: * sign. 5 %, ** sign. 1 %, ¹Latein, Griechisch oder Französisch

Die Korrelationen der LPS-Subtests und einzelner schulischer Leistungen, die Tent (1969) bei Schülern der 6. Klasse verschiedener Schularten (295 Jungen, Volksschule; 347 Mädchen, Volksschule; 103 Schüler, Realschule; 74 Schüler, Gymnasium) feststellte, sind in Tabelle 2.31 aufgeführt.

listet. Es wird deutlich, dass sowohl Persönlichkeitsvariablen wie Betragen, Fleiß, Aufmerksamkeit und Ordnung, als auch „Nebengegenstände“ (i. S. von Weiss, s. o.) wie Religion, Kunsterziehung, Musik und Sport nur vereinzelt in bedeutsamer Höhe mit einzelnen Subtests korrelieren. Für die Fächer Sozialkunde, Deutsch, Geschichte, Erdkunde, Biologie, Physik, Mathematik und Geometrie zeigen sich hingegen bedeutsame Korrelationen vor allem mit den Subtests 1, 2, 4, 5, 6 und 12 sowie dem Gesamtscore. Die anderen Subtests ohne sprachliche Komponente weisen hingegen kaum bedeutsame Zusammenhänge auf.

Marks, Seeber und Stümper (1975) haben in ihrer Untersuchung, in der sie die Faktorenstruktur unterschiedlicher Ausbildungsrichtungen verglichen, signifikante Validitätsangaben bezüglich eines Ausbildungskriteriums gemacht. Die Subtests 3 und 4 zeigen dabei teilweise höhere Zusammenhänge als der Gesamtscore. Insgesamt sind die Ergebnisse so zu bewerten, dass bivariat unterschiedlichste Zusammenhänge auszumachen sind und sich dadurch eine Eignung des Verfahrens für die Berufseignungsdiagnostik ableiten lässt.

Tabelle 2.30: Interkorrelationskoeffizienten zwischen den Untertests des LPS und Außenkriterien (nach Marks, Seeber & Stümper, 1975)

Kriterium	1	2	3	4	5	6	7	8	9	10	11	12	13	14	G
Lehrlinge DV, Facharbeiterabschluss	.20	.25	.19	.26					.19	.28					.25
Lehrlinge DV, Leistungseinschätzung	.19			.20						.23					.19
Lehrlinge Chemiefacharbeiter, theoretische Zensuren	.28	.22	.42	.36	.23	.29			.31	.23		.32	.40		.34
Lehrlinge Chemiefacharbeiter, Praxiseinschätzung	.31	.29	.46	.42	.27	.28		.27	.30	.28	.20	.32			.44
Elektromonteure, Einsatz zu Schalthandlungen	.21		.20				.22	.26		.24	.28	.14			
Kraftwerkmaschinenisten, Berufserfolg							.24	.31							
Chemieanlagenfahrer, Berufserfolg	.29		.25		.36	.23									
Locher, Normerfüllung					.21						.26				
Prüfer, Normerfüllung					.24										

Bemerkung: Es wurden in der Studie nur signifikante ($p < .05$) Korrelationskoeffizienten angegeben.

Tabelle 2.31: Interkorrelationskoeffizienten zwischen den Untertests des LPS und einzelnen Schulleistungen von Jungen und Mädchen der 6. Klasse Volksschule, sowie von Schülern der 6. Klasse Realschule und Gymnasium nicht nach Geschlecht getrennt (nach Tent, 1969)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	G
Be	.09 .05 .17 .26	.12 .04 .11 .20	-.01 .02 -.09 .08	-.03 .11 -.13 .11	-.01 -.01 .07 .20	-.06 -.02 -.11 .12	-.07 -.03 -.17 .01	-.03 .02 .11 -.02	.00 -.03 .07 -.09	.04 -.05 -.03 .00	-.09 -.02 -.16 -.09	-.07 .05 -.02 .17	-.02 .04 .07 -.06	-.10 .03 -.14 -.14	-.05 .02 -.08 .08
Fl	.19 .21 .00 .17	.05 .17 .16 .19	.05 .16 -.06 .05	.15 .27 .01 .10	.16 .21 .26 .19	.15 .16 .11 .16	.01 .00 .12 -.13	-.05 .08 .25 .14	.08 .03 .14 -.02	.13 .07 .11 .07	.10 .06 .12 -.09	.08 .15 .27 .07	.09 .13 -.01 .18	.00 .15 .24 -.13	.08 .20 .25 .13
Au	.16 .23 .00 .17	.13 .22 .14 .11	.05 .27 -.08 .01	.10 .32 -.14 .13	.20 .17 .19 .19	.13 .17 .10 .25	.02 .10 .09 -.13	.00 .14 .19 .14	.11 .18 .18 -.14	.08 .22 .07 .07	.04 .06 .04 .08	.09 .17 .12 .22	.06 .09 .03 .12	.00 .10 .13 .08	.08 .27 .18 .17
Or	.15 .19 -.16 .15	-.01 .12 .09 .03	.06 .18 -.06 -.08	.12 .25 -.10 .09	.14 .17 .20 .05	.09 .17 .14 .00	.01 .02 -.01 -.05	.04 .04 .07 .04	.06 .04 .02 -.13	.07 .08 .07 -.04	.03 .06 -.04 -.02	.09 .23 .05 .27	.09 .12 -.03 .06	-.02 .18 .11 .03	.04 .12 .08 .05
Re	.16 .26 .19 .15	.06 .21 .20 .15	.03 .14 -.01 -.09	.07 .24 -.01 -.04	.18 .18 .30 .22	.14 .25 .07 .27	.07 .01 -.01 -.04	.02 .12 .07 .13	.12 .15 .17 .01	.10 .12 .16 .13	.09 .18 .14 .11	.13 .20 .19 .12	.01 .15 -.01 .09	.07 .19 -.06 .10	.11 .26 .18 .18
So	.37 .32	.16 .28	.21 .32	.35 .30	.18 .18	.20 .23	.07 .00	.08 .16	.14 .20	.15 .23	.19 .15	.27 .24	.09 .18	.13 .16	.23 .34
Dm	.37 .42 .31 .50	.18 .38 .43 .42	.08 .22 .11 .16	.24 .33 .09 .20	.26 .30 .40 .44	.30 .31 .19 .43	.02 .03 .18 .02	-.02 .09 .21 .22	.16 .14 .13 -.04	.09 .15 .28 .12	.24 .19 .21 -.08	.28 .34 .45 .35	.08 .18 .06 .09	.10 .17 .17 .01	.18 .36 .42 .35
Ds	.36 .48 .31 .52	.25 .43 .40 .44	.13 .29 .09 .17	.29 .39 .15 .21	.30 .38 .40 .44	.26 .32 .28 .44	.00 .05 .07 .00	.01 .15 .12 .21	.10 .17 .07 -.05	.16 .15 .21 .14	.18 .17 .02 -.08	.37 .38 .34 .36	.10 .19 .10 .10	-.02 .17 .10 .00	.20 .40 .34 .35
Ge	.34 .40 .30 .23	.12 .32 .26 .14	.20 .34 .08 .00	.32 .39 .10 .17	.29 .29 .38 .26	.25 .24 .17 -.01	.03 -.03 -.01 -.01	.10 .19 .23 .19	.20 .22 .03 -.02	.19 .21 .26 .03	.22 .19 .19 .02	.20 .28 .28 .24	.08 .11 .01 .08	.09 .14 .08 .04	.22 .37 .32 .18
Er	.32 .36 .05 .38	.13 .31 .11 .30	.18 .32 .04 .09	.31 .39 .10 .23	.25 .27 .18 .20	.19 .28 -.01 .13	.11 .02 -.03 -.09	.04 .19 .15 .25	.25 .26 .12 .02	.16 .23 .22 .21	.15 .15 .05 .17	.24 .32 .16 .32	.11 .21 .08 .20	.16 .16 .11 -.05	.23 .38 .19 .30
Bi	.25 .34 .11 .16	.12 .29 .20 .11	.19 .30 .03 -.02	.32 .33 .10 .24	.25 .23 .18 .03	.18 .31 .17 -.01	.04 .09 .13 -.05	.09 .16 .13 .20	.21 .22 .13 .08	.16 .20 .31 .20	.20 .21 .14 .10	.20 .21 .24 .20	.10 .16 .00 .08	.02 .15 .24 .03	.18 .35 .28 .18
Ph	.28 .32	.11 .26	.18 .32	.31 .37	.21 .16	.20 .21	.04 .07	.10 .22	.18 .26	.13 .22	.15 .16	.14 .20	.04 .15	.05 .17	.17 .35
Ma	.30 .25 .02 .29	.15 .23 .18 .24	.25 .38 .14 .25	.45 .46 .17 .44	.23 .25 .10 .31	.22 .24 .07 .08	-.02 .04 .11 .19	.18 .22 .34 .23	.25 .17 .13 .19	.15 .27 .18 .03	.11 .08 .19 .01	.18 .25 .04 .22	.08 .15 .01 .34	.08 .12 .22 .05	.20 .37 .24 .34
Ra	.33 .26	.19 .28	.31 .41	.44 .45	.24 .22	.21 .30	.01 .00	.13 .27	.22 .25	.11 .29	.12 .08	.23 .26	.10 .16	-.01 .15	.18 .42
Mu	.16 .18 .16 .11	.18 .13 .11 .16	.00 .26 .13 -.02	.12 .32 .13 -.01	.19 .11 .13 .14	-.07 .11 .08 .28	-.08 .00 -.05 -.10	.03 .14 .02 -.06	.02 .15 .14 -.19	-.05 .18 .08 .12	.10 .12 .16 .13	.05 .14 .07 .11	.02 .11 .08 .13	-.02 .10 -.05 .30	.08 .26 .03 .10
Ku	.17 .18 -.08 .03	.07 .12 -.06 -.13	.04 .13 -.12 .04	.23 .18 .03 .05	.19 .18 .07 -.06	.18 .09 -.01 .03	.00 .00 -.11 -.01	.00 .02 .12 .05	.12 .10 -.06 -.02	.08 .04 .05 .20	.11 .14 .01 .22	.16 .18 -.04 .18	.12 .15 .04 .14	-.01 .06 -.08 .26	.12 .18 .00 .13
Sp	.07 .12 .04 .16	.12 .11 .03 .03	-.05 .16 -.02 -.17	.04 .17 -.04 -.18	.11 .14 -.06 .05	.11 .08 -.07 .24	.05 .03 .15 .08	-.04 .13 -.06 .08	-.01 .10 .04 .14	-.01 .10 .03 .01	.06 .18 .08 .16	.07 .10 -.01 .03	.09 .15 .03 .16	.10 .10 .14 .17	.04 .18 .03 .13
Fr	.10 .41	.24 .32	.04 .02	.06 .15	.15 .32	.02 .25	.02 .01	.04 .23	-.03 -.18	.07 .20	.07 .08	.12 .50	-.11 .15	.21 .09	.13 .32
Ha	.23 .15 -.08 .33	.15 .14 .04 .31	.03 .10 .01 .16	.17 .18 -.12 -.02	.16 .12 .19 .32	.09 .20 .05 .35	.05 .06 -.02 .06	.02 .02 -.01 .11	.04 .09 -.08 -.04	.11 .00 .09 .38	.11 .07 -.01 -.01	.23 .28 .10 .21	.08 .12 -.02 .17	.03 .15 .06 .11	.13 .22 .04 .31

Bemerkungen: Der Aufbau der einzelnen Felder entspricht dem Schema

Jungen Volksschule	Mädchen Volksschule
Realschule	Gymnasium

Zwar folgt dies keiner strengen Alphaadjustierung, aber als Signifikanzkriterium wird das stärkste für diese vier Stichproben angegeben: 5 %-Niveau: $r > .23$, 1 % Niveau: $r > .30$

Be – Betragen; Fl – Fleiß; Au – Aufmerksamkeit; Or – Ordnung; Re – Religion; So – Sozialkunde; Dm – Deutsch mündlich; Ds – Deutsch schriftlich; Ge – Geschichte; Er – Erdkunde; Bi – Biologie, Naturkunde; Ph – Physik, Naturlehre; Ma – Mathematik, Rechnen; Ra – Raumlehre, Geometrie; Ku – Kunsterziehung; Mu – Musik; Sp – Sport, Leibesübung; Fr – Fremdsprache (Englisch, Französisch, Latein); Ha – Handschrift

2.4 Diskussion

Die vorgestellten Ergebnisse legen den Schluss nahe, dass eine Reduktion der Zahl der Subtests des LPS und somit ein effizienteres und vielleicht auch aussagekräftigeres Diagnoseinstrument möglich ist. Die häufige Zusammenlegung von Subtest 1 und 2 zeigt deutlich, dass eine gesonderte Interpretation der Ergebnisse nicht gewünscht wird. Die hohen Ladungen der beiden einzelnen Subtests auf immer einen Faktor weisen ebenfalls darauf hin, dass eine Trennung nicht nötig ist. Die Konsequenzen der Testverkürzung können bei den gegebenen hohen Reliabilitätskennwerten beider Tests in Hinblick auf die Praktikabilität und Ökonomie akzeptiert werden. Ferner gilt es, die Zusammenhänge der weiteren Subtests für sprachliche Fähigkeiten 5, 6 und 12 zu überdenken. Für die Revision des PSB (Horn, PSB 6-13 R, 2003) wurde das Aufgabenprinzip von Subtest 6 beibehalten, von Subtest 12 hingegen nicht. Alle drei Subtests weisen ähnliche Korrelationen mit anderen Tests und externen Kriterien auf. Am häufigsten korreliert Subtest 12 mit allen IST-Subtests und auch anderen Intelligenz- und Leistungstests. Auch die hohe Korrelation von Subtest 12 mit dem LPS-Gesamtwert (.88, Horn, 1983) deutet darauf hin, dass mit diesem Subtest nicht allein Closure sondern auch andere Primary Mental Abilities erhoben werden. Da Subtest 6 für eine Anwendung am Computer wenig geeignet ist - die Eingabe von Wörtern über die Tastatur oder auch das Schreiben auf einem Tablet-PC würde große bzw. noch ungeprüfte Schreibgeschwindigkeitsvarianzen erzeugen - erscheint dieser Test im Sinne der Ökonomie und auch der Vergleichbarkeit bei einer steigenden Relevanz der Testung am Computer als nicht mehr in das Leistungsprüfsystem integrierbar. Die hohe Ähnlichkeit der statistischen Zusammenhänge der Subtests 5 und 12 mit den anderen Subtests sowie anderen Kriterien verleitet zur Idee, das Prinzip der Verstümmelung der Worte zur Erhebung von Closure (Subtest 12) auf das Prinzip des Worterkennens und das Ankreuzen des ersten Buchstabens eines Wortes (Subtest 5) anzuwenden, da das Konstrukt Wordfluency offensichtlich nicht in geeignetem Maße von Verbal Comprehension getrennt werden konnte, was nach Thurstone (1969) nur mit einer großen Anzahl von Tests möglich ist.

Die Subtests 3 und 4 laden sehr häufig auf einem gemeinsamen Faktor. Das Reasoning-Konstrukt scheint durch diese Subtests gut erfasst zu werden. Die Revision des PSB, in der zwischen Zahlen-, Buchstaben- und Figurenreihen unterschieden wird, zeigt aber ebenso wie die Kritik von Steller und Stürmer (1984), dass Reasoning und Space nicht ohne Weiteres getrennt werden können. Sowohl in der Literatur als auch in der Reanalyse laden die Subtests 3 und 4 nicht unwesentlich häufig zusammen mit mindestens einem der Subtests 7 bis 10 auf einem Faktor. In der Studie von Weiss (1972) erweist sich Subtest 4 als stärkerer Prädiktor für Schulleistungen als Subtest 3. Die Revision des PSB legt nahe, dass durch die Trennung der drei Formen in besonderen Fällen spezifischere Aussagen getroffen werden können, eine Messung des Konstrukts Reasoning aber besser durch die Zusammenlegung aller drei Formen möglich ist.

Die Zusammenfassung der Subtests 7 bis 10 erscheint - ohne Diskussion der Validität - statistisch als sinnvoll bezeichnet werden zu können, da alle vier Subtests relativ häufig auf gemeinsamen Faktoren laden. Allerdings ist hierbei zu berücksichtigen, dass dies teilweise unabhängig voneinander geschieht. Der Vergleich der Literatur mit der Reanalyse zeigt, dass Subtest 7 vornehmlich durch das strengere Kriterium mit den anderen dreien, aber auch mit einzelnen auf einem Faktor lädt. Subtest 8 ist seit längerem umstritten und wurde beispielsweise in die Versi-

on LPS 50+ nicht mehr aufgenommen (Sturm, Willmes & Horn, 1993). Er lädt zudem sehr häufig mit den Subtests 9 und 10 auf gemeinsame Faktoren. Ebenso laden die Subtests 8 und 10 in den Analysen des PSB immer auf demselben Faktor. Die Ergebnisse bezüglich der Subtests 9 und 10 sind deskriptiv relativ ähnlich. Beide Subtests laden häufig auf demselben Faktor. Die Korrelationen mit einzelnen Kriterien sind bei allen vier Subtests ähnlich. Bei größeren Stichproben sind fast alle Korrelationen mit den Subtests anderer Verfahren signifikant. Bezüglich der Schulnoten zeigen sich nur vereinzelt bedeutsame Korrelationen. Für die Leistung der Lehrlinge sind eher die Subtests 9 und 10, für die Leistung von Monteuren und Maschinisten eher die Subtests 7 und 8 sensibel (Marks et al., 1975, Tabelle 2.30). Der Vergleich der Subtests 3, 4, 8, 9 und 10 mit der Testbatterie des Berufsförderungswerks Eckert (Berufsinteressentest II [BIT II, Irle & Allehoff, 1984], Intelligenz-Struktur-Test 2000 R [IST 2000 R, Amthauer, Brocke, Liepmann & Beauducel, 2001], FRT, Mechanisch-Technischen Verständnistest [MTVT, Lienert, 1958], Mannheimer Rechtschreibtest [MRT, Jäger & Jundt, 1981], d2, KLT, Freiburger Persönlichkeitsinventar [FPI-R, Fahrenberg, Hampel & Selg, 2001], Berufsbezogenen Rechentest [BRT, Balser, Ringsdorfer & Traxler, 1986], ein Erinnerungstest und mehrere Algebratests, die vom Psychologischen Dienst selbst konzipiert wurden; Kreuzpointner, 2008) zeigt in einer über alle Tests hinweg durchgeführten Faktorenanalyse, dass die Subtests 3, 4, 9 und 10 des LPS mit dem MTVT auf einem, Subtest 8 mit dem FRT, dem IST FA und dem IST WÜ auf einem andern Faktor laden. Die Korrelationen der Subtests 8, 9 und 10 mit den anderen Testwerten unterscheiden sich nur unwesentlich. Subtest 11 zeigt über alle Studien hinweg nicht den erwarteten Zusammenhang zu Subtest 12. Dies liegt entweder an dessen starken verbalen Einflüssen oder an der Tatsache, dass Speed of Closure in der gegebenen Form nicht erhoben werden kann. Die häufigen gemeinsamen hohen Ladungen von Subtest 11 mit den Subtests 7 bis 10 legen für die Testleistung bei Subtest 11 einen starken Einfluss der räumlichen Wahrnehmung nahe.

Die Subtests 13, 14 und 15 liegen häufig auf einem Faktor, vor allem in den Studien, die Subtest 15 anwendeten. In Studien ohne Subtest 15 laden 13 und 14 seltener auf demselben Faktor. Es wird deutlich, dass für Subtest 13 und 14 der Umgang mit Zahlen, die hier im Grunde eigentlich nur Mittel zum Zweck sind, einen entscheidenden Einfluss hat, da eine schnelle und richtige Leistung bei der Addition in Subtest 15 in beiden Untertests eine Rolle spielt. Subtest 13 und 14 haben insgesamt die geringsten Korrelationen mit den Subtests anderer Intelligenztests und kaum signifikante Korrelationen mit Schulleistungen. Im PSB wird deutlich, dass Subtest 14 eine von den anderen Subtests unterschiedliche Fähigkeit misst.

2.5 Schlussfolgerung und Revisionsvorschlag

Aufgrund der vorgestellten Ergebnisse und Interpretationen wird nun eine neue Zusammenstellung des Leistungsprüfsystems vorgeschlagen und im Anschluss abschließend begründet.

Tabelle 2.32: Bearbeitungsdauer der Subtests und Vorschlag für die Neuzusammenstellung des LPS

LPS-neu	Entsprechungen des LPS (Horn, 1962, 1983)	Bearbeitungsdauer LPS	Geschätzte Bearbeitungs- dauer LPS-neu
1	1 + 2	2' + 3'	3'
2	12	2'	2'
3	3	5'	4'
4	4 Zahlen	8'	4'
5	4 Buchstaben		4'
6	7	2'	2'
7	9	3'	3'
8	10	3'	3'
9	13	8'	5'
10	14	2'	2'
11	15	8'	5'
	5, 6, 8, 11	3' + 1' + 4' + 1'	
Nettobearbeitungsdauer		55'	37'

Bemerkung: LPS-neu - Revisionsvorschlag

Diese neue Zusammenstellung folgt drei Prinzipien: 1. Erhalt des Leitgedankens des LPS, 2. Erhöhung der Ökonomie und Praktikabilität und 3. Neuorientierung der strukturtheoretischen Grundlagen. Die Revision eines Diagnoseverfahrens darf die Prinzipien, den Aufbau und die Form des Originals nicht in dem Maße verändern oder verzerren, so dass es sich um einen neuen Test mit altem Namen handelt. Horn (1962, 1983) erarbeitete mit dem LPS ein Verfahren, das vor allem in der Berufseignungsdiagnostik zum Einsatz kommen und für alle Schularten geeignet sein sollte. Bei der Entstehung des LPS waren die Primary Mental Abilities von Thurstone (1938) maßgeblich richtungweisend. In der Überarbeitung (Horn, 1983) wird zudem ein eigenes, auf einer Pyramide basierendes Intelligenzmodell vorgestellt (vgl. Abbildung 2.1).

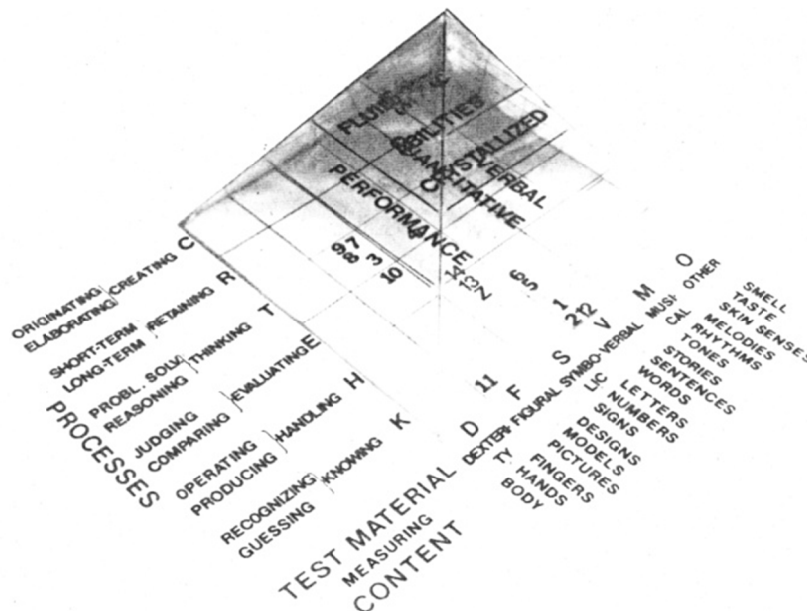


Abbildung 2.1: Intelligenzpyramide nach Horn (1983, S. 30)

3. Das neue Leistungsprüfsystem (LPS-neu)

Im Folgenden wird das Resultat der Analysen der in Kapitel 2 dargestellten Studien über das Leistungsprüfsystem von Horn (1983) vorgestellt. Die einzelnen Items werden in dieser Arbeit nicht dargestellt, da sie zum einem dem bisherigen LPS entnommen wurden und zum anderen in der Revision des LPS Verwendung finden sollen. Eine Veröffentlichung der Items außerhalb des Manuals ist jedoch nicht im Sinne eines Leistungstests. Gleichwohl werden die Instruktionen und die darin verwendeten Beispiele dargestellt, so dass nachvollziehbar wird, wie die einzelnen Subtests aufgebaut sind.

Beim Aufbau der Papier-Bleistift-Version (PBV) des LPS-neu wurde versucht, den Prinzipien des LPS von Horn so weit wie möglich treu zu bleiben. Insbesondere die Bearbeitung der Items direkt in der Vorgabe sollte beibehalten werden. Ebenso die Konzeption des Testbogens auf vier Seiten, so dass der Druck auf einem DIN A 3 Bogen möglich ist (Der originale LPS-Bogen ist ein wenig kleiner als DIN A 3.). Der daraus resultierenden relativ gedrängten Darstellung der Items, die von einigen Seiten derart kritisiert wurde, das LPS teste mehr die Sehfähigkeit als kognitive Leistungen, wurde durch verschiedene Maßnahmen begegnet. Zum einen ermöglichten die Digitalisierung der Items und der Druck auf weißes Papier einen deutlich größeren Kontrast. Zudem wurden die Antwortmodalitäten bei einzelnen Subtests zugunsten der Übersichtlichkeit geändert, so dass die Items an sich besser zu erkennen sind. Des Weiteren wurden die Instruktionen und Itembeispiele nicht mehr an den Anfang einer Itemkolonne gestellt, sondern auf der ersten Seite des Testbogens zusammengefasst. Dies und der Raumgewinn durch das etwas größere Papierformat ermöglichten eine Darstellung der Items, die bis dato bei kaum einer der Testpersonen zu Beschwerden bezüglich der Erkennbarkeit der Aufgaben geführt hat. In Anhang 15.1 werden die Seiten des Testbogens verkleinert dargestellt. Zum Schutz der Items werden nur die Beispielitems als Platzhalter für die übrigen Items benutzt.

3.1 Die Subtests des revidierten Leistungsprüfsystems

Das neue LPS besteht, wie bereits in Tabelle 2.32 dargestellt, aus elf Subtests. Subtest 1 ist aus den Items der Subtests 1 und 2 des alten LPS zusammengestellt. Die Trennung in zwei Subtests wurde inhaltlich von Horn (1962; 1983) nicht begründet und hatte zur Folge, dass die Ergebnisse der beiden Untertests in der Regel immer gemittelt interpretiert wurden. Die hohe Korrelation der beiden Subtests belegt ihre Übereinstimmung. Der messtheoretische Gewinn an Genauigkeit der Messung durch die doppelte Testung soll durch die Erhöhung der Itemzahl aufgrund von Zusammenlegung und erster Aussortierung von als ungeeignet erscheinenden Items erhalten bleiben. Mit Subtest 1 soll Allgemeinbildung gemessen werden. In den vorgegebenen Wörtern wird jeweils ein Buchstabe verändert. Die Aufgabe des Probanden besteht darin, diesen falschen Buchstaben zu identifizieren. Die Idee dieser Testung ist, dass Probanden mit einer höheren Allgemeinbildung zum einen die Worte mit einer höheren Wahrscheinlichkeit kennen und zum anderen deren Orthographie wissen, so dass es ihnen eher möglich ist, den Fehler zu finden. In Abbildung 3.1 wird der Teil der Instruktionsseite abgebildet, der sich auf die Subtests 1 und 2 beziehen. Die ausführliche Instruktion wird durch den Versuchsleiter vorgelesen oder auswen-

dig vorgetragen. Sie basiert im Wesentlichen auf der ursprünglichen Form, wurde jedoch den aktuellen Gegebenheiten angepasst (siehe Anhang 15.2).

<div style="border: 1px solid black; border-radius: 50%; width: 40px; height: 40px; margin: 0 auto; display: flex; align-items: center; justify-content: center; margin-bottom: 10px;">1</div> <p style="text-align: center;">Markieren Sie den falschen Buchstaben.</p> <p style="text-align: center; font-size: 1.2em; letter-spacing: 0.5em;">K R X I D E</p> <p style="font-size: 0.8em;">Um eine falsche Auswahl aufzuheben, zeichnen Sie einen Kreis darum und kreuzen Sie stattdessen ihre alternative Lösung an.</p> <p style="text-align: center; font-size: 1.2em; letter-spacing: 0.5em;">T E X L X R</p>	<div style="border: 1px solid black; border-radius: 50%; width: 40px; height: 40px; margin: 0 auto; display: flex; align-items: center; justify-content: center; margin-bottom: 10px;">2</div> <p style="text-align: center;">Markieren Sie den Anfangsbuchstaben.</p> <p style="text-align: center; font-size: 1.2em; letter-spacing: 0.5em;">G X W E R</p> <p style="font-size: 0.8em;">2. Beispiel:</p> <p style="text-align: center; font-size: 1.2em; letter-spacing: 0.5em;">C K E R X</p>
--	---

Abbildung 3.1: Instruktionen der Subtests 1 und 2 aus dem Aufgabenbogen des LPS-neu

Subtest 2 des LPS-neu stellt eine Kombination aus den Subtests 5 (Anagramme) und 12 (Verstümmelung der Darstellung der Wörter) der alten Version dar. Wie in Kapitel 2.2 und insbesondere in Tabelle 2.23 deutlich wird, hängen beide Subtests besonders mit der Allgemeinbildung, die über die sprachliche Kompetenz abgebildet wird, zusammen. Auch hierbei wird davon ausgegangen, dass ein erhöhtes allgemeines Wissen dazu führt, dass die Wörter, diesmal als verstümmelte Anagramme dargestellt, eher erkannt werden. Aus den Vorgaben des alten LPS wurde die relativ geringe Veränderung der Buchstabenreihenfolge der einzelnen Items übernommen. Falls sich die Items dadurch als zu leicht lösbar erweisen sollten, besteht für weitere Fassungen die Möglichkeit, die Reihenfolge der Buchstaben stärker zu mischen.

Subtest 3 wurde vollkommen identisch aus dem alten LPS übernommen. Gemeinsam mit den Subtests 4 und 5 soll damit Reasoning im Sinne der Primary Mental Abilities (Thurstone, 1938) erhoben werden (vgl. Abbildung 3.2). Basierend auf den Erkenntnissen der Revision des PSB (Horn, 2003) wurden in den Subtests 4 und 5 Items mit Zahlen und Items mit Buchstaben getrennt. In allen drei Aufgabebetypen ist es die Aufgabe des Probanden, die Regel, die der Systematik der acht bzw. neun Symbole zugrunde liegt, zu identifizieren und dasjenige Symbol zu markieren, das dieser Regel zuwider läuft. Anastasi (1976) definiert (General) Reasoning bzw. „Induction“, also deduktives und induktives Denken, wie folgt:

Induction (or General Reasoning): *The identification of this factor was least clear. Thurstone originally proposed an inductive and a deductive factor. The latter was best measured by tests of syllogistic reasoning and the former by tests requiring the subject to find a rule, as in a number series completion test. Evidence for the deductive factor however, was much weaker than for the inductive. Moreover, other investigators suggested a general reasoning factor, best measured by arithmetic reasoning tests. (Anastasi, 1976, S. 373)*

Durch die Aufgabengestaltung wird versucht, beiden Arten des Schließens gerecht zu werden. Die erste Aufgabe des Probanden ist es, die zugrunde liegende Regel zu identifizieren. Dies entspricht dem induktiven Faktor. Unter der Bedingung der Gültigkeit der entdeckten Regel muss dann das Symbol gefunden werden, das nicht zur Regel passt. Hierbei können deduktive Prozesse angenommen werden. Insgesamt soll also der Gedanke des General Reasoning Faktors aufge-

griffen und abgebildet werden. Aufgrund der Darstellungen im Manual des alten LPS wird davon ausgegangen, dass dies im Sinne Horns ist.

<div style="border: 1px solid black; border-radius: 50%; width: 40px; height: 40px; margin: 0 auto; display: flex; align-items: center; justify-content: center; margin-bottom: 10px;">3</div> <p>Markieren Sie die Form, die nicht in die Reihe passt.</p> <div style="display: flex; justify-content: space-around; align-items: center; margin: 10px 0;"> ○○○○○○○○ </div> <p>Es passt immer nur eine Form nicht in die Reihe!</p> <div style="display: flex; justify-content: space-around; align-items: center;"> + + + ✕ </div>	<div style="border: 1px solid black; border-radius: 50%; width: 40px; height: 40px; margin: 0 auto; display: flex; align-items: center; justify-content: center; margin-bottom: 10px;">4</div> <p>Markieren Sie die Zahl, die nicht in die Reihe passt.</p> <div style="display: flex; justify-content: space-around; align-items: center; margin: 10px 0;"> 222222222 </div> <p>Es passt immer nur eine Zahl nicht in die Reihe!</p> <div style="display: flex; justify-content: space-around; align-items: center;"> 455445445 </div>	<div style="border: 1px solid black; border-radius: 50%; width: 40px; height: 40px; margin: 0 auto; display: flex; align-items: center; justify-content: center; margin-bottom: 10px;">5</div> <p>Markieren Sie den Buchstaben bzw. das Buchstabenpaar, das nicht in die Reihe passt.</p> <div style="display: flex; justify-content: space-around; align-items: center; margin: 10px 0;"> abababaaa </div> <p>Es passt immer nur ein Buchstabe bzw. ein Buchstabenpaar nicht in die Reihe!</p> <div style="display: flex; justify-content: space-around; align-items: center;"> aabbbaaaabaaaab </div>
---	---	---

Abbildung 3.2: Instruktionen der Subtests 3, 4 und 5 aus dem Aufgabenbogen des neuen LPS

Die Items der Subtests 6 bis 10 wurden ebenfalls aus dem alten LPS übernommen (vgl. Abbildung 3.3 und Abbildung 3.4). Für die Subtests 7 und 8 wurde die Antwortmodalität ein wenig verändert. Für Subtest 11 wurden die ersten 80 Items übernommen. Subtest 6 (7 in der alten Version) dient zur Ermittlung des Faktors Space. Eine Zahl oder ein Buchstabe pro Zeile ist spiegelverkehrt abgebildet. Zudem sind alle Zeichen um den Mittelpunkt rotiert. Die Aufgabe besteht nun darin, durch eine mentale Rotation und den Vergleich der Symbole das spiegelverkehrte Zeichen zu identifizieren. Somit müssen die beiden Faktoren, die Space definieren, genutzt und können so erhoben werden:

Space: *This factor may represent two distinct factors, one covering perception of fixed spatial or geometric relations, the other manipulatory visualizations, in which changed positions or transformations must visualized. (Anastasi, 1976, S. 372)*

Ebenso müssen die geometrischen Figuren in Subtest 7 (9 in der alten Version) bei der Aufgabenlösung, die Anzahl der Flächen zu bestimmen, nun im dreidimensionalen Raum visualisiert werden. Es kann zudem davon ausgegangen werden, dass insbesondere bei komplexeren Figuren als den Beispielen, zur Visualisierung auch Rotationen durchgeführt werden, um die Zahl der Flächen zu bestimmen. Im Unterschied zum alten LPS werden für jede Figur die Zahlen 2 bis 10 als Antwort vorgegeben, so dass das Antwortformat bei allen Items gleich ist und somit die Objektivität erhöht wird.

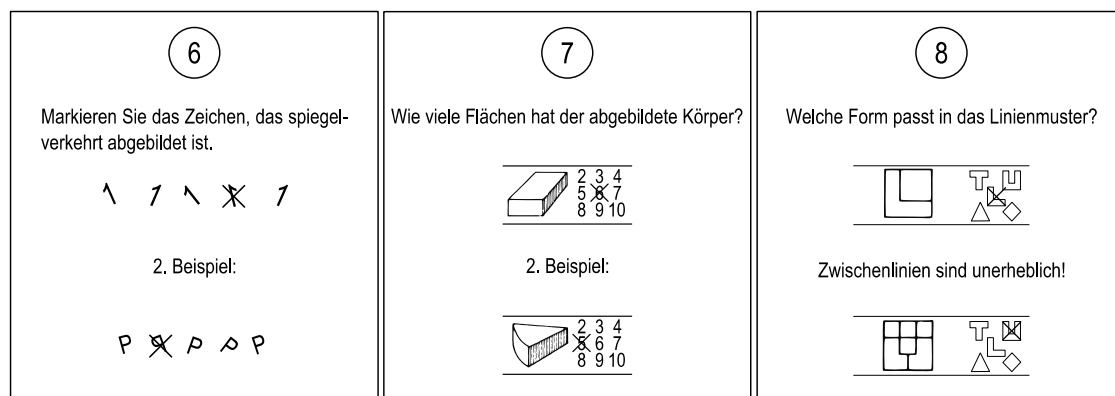


Abbildung 3.3: Instruktionen der Subtests 6, 7 und 8 aus dem Aufgabenbogen des LPS-neu

In Subtest 8 (10 im alten LPS) wird keine mentale Rotation von den Probanden gefordert. Die Aufgabe besteht darin, einen der fünf Umrisse, die neben jedem Muster abgebildet sind, in diesem zu erkennen. Hierfür müssen die Umrisse mental etwas vergrößert und anschließend über das Muster projiziert werden. Es ist auch vorstellbar, dass Repräsentationen der Umrisse – schließlich stellen sie bekannte Formen dar – durch mentales Nachzeichnen zu bestätigen versucht werden.

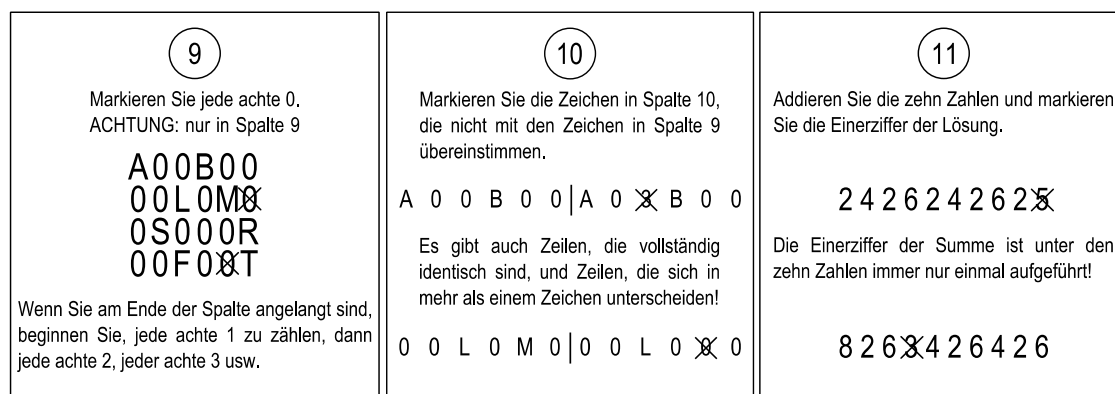


Abbildung 3.4: Instruktionen der Subtests 9, 10 und 11 aus dem Aufgabenbogen des LPS-neu

Wie im alten LPS wird Subtest 10 (alt: 14) vor Subtest 9 (alt: 13) durchgeführt. Die Spalte der Items von Subtest 10 stellt eine Kopie der Spalte von Subtest 9 dar, wobei bei 40 der 60 Zeilen mindestens ein Zeichen verändert wurde. Die Aufgabe des Probanden ist es, diese veränderten Zeichen zu identifizieren. Ab der 31. Zeile bestehen sie aus je acht und nicht mehr aus nur sechs Zeichen. Während der Testung wurde deutlich, dass Linkshänder eine gewisse Benachteiligung aufgrund der Anordnung haben, da sie sich teilweise mit der Schreibhand über die zu vergleichenden Zeichen bewegten. Durch den Hinweis, eine für den Vergleich vorteilhaftere Haltung einzunehmen, wurde versucht, diesen Nachteil auszugleichen. Inwiefern dies eine relevante Störgröße darstellt, wird nicht beurteilt werden können, da die Händigkeit nicht erhoben wurde. Diese Entscheidung muss auf spätere Untersuchungen verschoben werden. Die Subtests 9 und 10 erheben „Perceptual Speed“:

Perceptual Speed: *Quick and accurate grasping of visual details, similarities, and differences. This factor may be the same as the speed factor identified by Kelley and other early investigators. This is one of several factors subsequently identified in perceptual tasks (Thurstone, 1944).* (Anastasi, 1976, S. 372)

In Subtest 10 geht es also darum, den zeilenweisen Vergleich so schnell wie möglich durchzuführen. Die Gefahr von Fehlern, insbesondere von falsch-positiven, also der Identifikation von Unterschieden, wo keine sind, ist relativ gering und zudem im Vergleich zu Subtest 9 mit relativ geringen Konsequenzen bezüglich der Rohwertsumme verbunden. In Subtest 9 besteht die Aufgabe des Probanden darin, im ersten Durchgang jede achte „0“ zu markieren. Am Ende der Spalte angelangt soll der Proband wieder von vorne beginnen und jede achte „1“ markieren usw. Es wird deutlich, dass ein früherer Fehler (z. B. wenn die 15. anstatt der 16. „0“ markiert wird) die Konsequenz haben kann, dass alle folgenden „0“, auch wenn sie acht „0“ von der vorhergehend markierten entfernt sind, als falsch gewertet werden. In Subtest 9 ist somit die Sorgfältigkeit neben der Schnelligkeit ein wesentliches Merkmal, das zu einer hohen Punktzahl führt.

Subtest 11, der als Subtest 15 im alten LPS nur als Zusatz geführt wurde, dient der Erhebung der Konzentrationsleistungsfähigkeit. Es wird davon ausgegangen, dass die Fähigkeit, Ziffern zu addieren, bei allen Probanden gegeben ist, so dass die Aufgabe, zehn einziffrige Zahlen zu summieren, von jedem erfüllt werden kann. Den leistungsbeeinflussenden Faktor stellt dann die Informationsverarbeitungsgeschwindigkeit dar: Wie schnell kann diese basale Fähigkeit, zu einer Zahl eine Ziffer zu addieren, fehlerfrei ausgeführt werden.

3.2 Das Intelligenzmodell des LPS-neu

Die theoretische Fundierung des alten LPS, insbesondere dessen Abbildung der Primary Mental Abilities, wird seit jeher als kritisch betrachtet (z. B. Greif, 1972; Lutz, 1983). 1983 beschreibt Horn ein Modell der Begabungsstruktur (Abbildung 2.1 und Abbildung 2.2), zu dem sich jedoch keine weiteren Veröffentlichungen finden lassen. Die darin dargestellte Unterteilung der Tests in verschiedene Arten von Aufgabeninhalten und unterschiedliche Operationen der Aufgabenbeantwortung stellt in seiner Komplexität eine Mischung aus dem Intelligenz-Struktur-Modell von Guilford (z. B. Guilford & Hoepfner, 1976) und dem Berliner Modell von Jäger (1982; Jäger, Süß & Beauducel, 1997) dar. Bei genauerer Betrachtung fällt auf, dass das LPS diesem Modell nicht gerecht wird, da die Subtests nur die Hälfte der proklamierten Kombinationen von Inhalten und Operationen abdecken. Der Berliner Intelligenzstruktur-Test (Jäger, Süß & Beauducel, 1997), der in Interaktion mit dem Berliner Modell zu dessen Validierung entwickelt wurde, bietet beispielsweise für jede der Kombinationen mehrere Subtests. Das Verfahren ist dafür wesentlich aufwändiger und zeitintensiver.

Insgesamt erscheint das Drei-Ebenen-Modell von Carroll (1993; 2005; The Three-Stratum Theory; vgl. Abbildung 3.5) wesentlich adäquater als Grundlage für das bereits bestehende Leistungsprüfsystem. Im Modell wird auf der obersten Ebene ein *g*-Faktor, also eine Allgemeine Intelligenz, die jegliche intellektuelle bzw. kognitive Leistung mehr oder weniger stark beeinflusst, postuliert. Die zweite Ebene besteht aus acht (wobei diese Zahl nicht als endgültig angesehen wird) Faktoren, die als „broad“, also in etwa als „allgemein“, bezeichnet werden. Diese Faktoren

liegen auf einem etwas höheren Abstraktionsniveau als die Primary Mental Abilities nach Thurstone (1938). Die unterste Ebene stellt umschriebene Fähigkeiten dar, die durch einen oder mehrere Test erhoben werden können. Dieses Stratum I, das „native performance“ beschreibt, divergiert im Abstraktionsniveau zwischen relativ konkreten Fertigkeiten (z. B. Cloze Ability, also die Fertigkeit, Lückentexte zu bearbeiten) und immer noch umfassenderen Konzepten wie z. B. Induction.

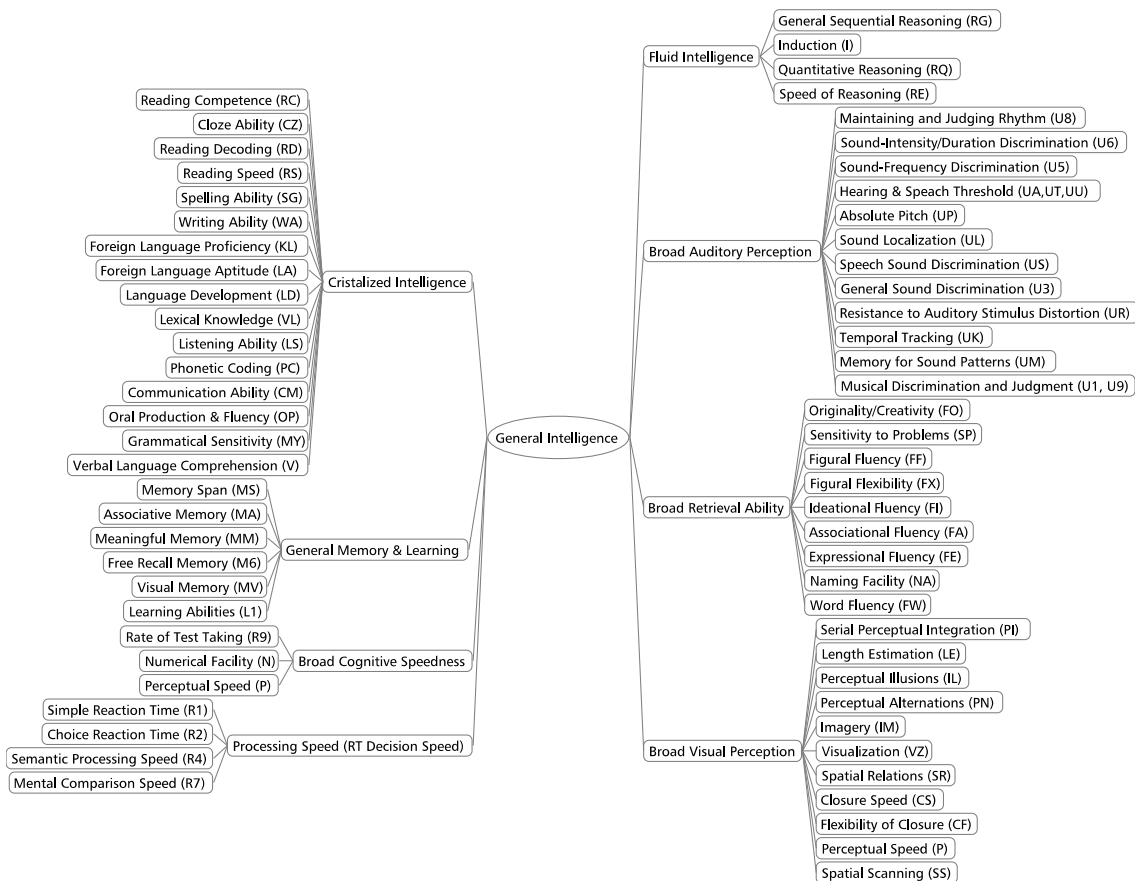


Abbildung 3.5: Die „three-stratum structure of cognitive abilities“ (nach Carroll, 1993, S. 626)

Der jahrzehntelange Streit zwischen Spearman und Thurstone, ob es nun einen *g*-Faktor und unbestimmt viele spezifische Fähigkeiten oder keinen generellen Faktor, sondern lediglich eine bestimmte Anzahl (sieben bis neun) von statistisch unabhängigen also unkorrelierten Fähigkeiten gebe, wurde schließlich durch Zugeständnisse auf beiden Seiten (Spearman & Jones, 1950 bzw. Thurstone & Thurstone, 1941) beendet (McGrew, 2005). Diese und viele andere Forscher im Bereich der Intelligenztestung gehen also davon aus, dass sich intellektuelle Fähigkeiten durch einen übergeordneten Faktor (Allgemeine Intelligenz) und weitere nicht generelle Faktoren, die durch die einzelnen Tests definiert werden, abbilden lassen. Carroll (1993) sammelte 477 brauchbare Korrelationsmatrizen von umfassenderen Intelligenztestbatterien und reanalyisierte dieses Datenmaterial jeweils mittels desselben Algorithmus (Faktorenanalyse nach Schmid & Leimann, 1957, die Faktoren höherer Ordnung zulässt). 2003 versuchte Carroll in einer Denkschrift für Arthur Vernon darzulegen, dass ein Modell ohne *g*-Faktor, insbesondere das Modell von Cattell und Horn (z. B. Horn & Blankson, 2005), für die Modellierung intellektueller

Fähigkeiten nicht adäquat ist. Insbesondere die Interkorrelationen von als unabhängig konstruierten Faktoren sieht Carroll als Beleg für die Notwendigkeit eines Faktors höherer Ordnung. In der Beschreibung des Modells stellt Carroll (2005) aber klar, dass ein so umfassendes Modell (vgl. Abbildung 3.5) nur höchst selten in seiner Gänze überprüft werden kann, und fordert daher eine Vielzahl von Untersuchungen, die sich den jeweiligen Teilaspekten widmet. Im Folgenden soll nun das LPS-neu in das Drei-Ebenen-Modell eingebettet werden.

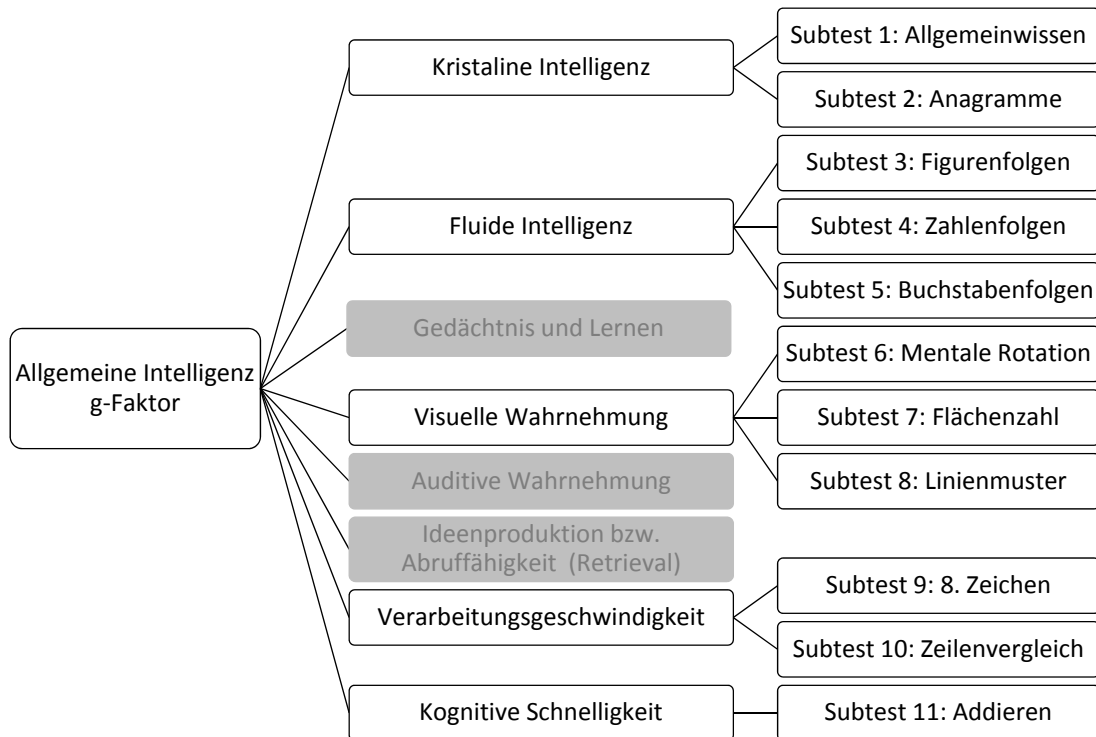


Abbildung 3.6: Vorschlag für die Zuordnung des LPS-neu in die Struktur des Drei-Ebenen-Modells nach Carroll (1993)

Es ist nicht das Ziel dieser Arbeit, die exakten Entsprechungen der einzelnen Subtests zu den von Carroll formulierten Stratum I Fähigkeiten zu überprüfen. Es soll jedoch die Einordnung der durch die Messung mittels der Subtests postulierten Fähigkeiten in die Stratum II Ebene vollzogen werden, indem diese mit ähnlich erscheinenden Fähigkeiten, deren theoriekonforme Zuordnung bereits bei Carroll diskutiert wurde, verglichen werden, um so eine Basis für den postulierten Zusammenhang und die Überprüfung dieses Modells (Abbildung 3.6) zu haben. Carroll (1993, S. 157) beschreibt einen etwas allgemeineren Verbalen Faktor (Verbal or printed language Comprehension, V): „In general, factors were classified as V when (a) all or a majority of their variables involved printed tests requiring reading, and (b) the variables covered a wide range of test types measuring general language development including (typically) various types of vocabulary tests and reading comprehension tests.“ Dieser begründet neben anderen Faktoren die Kristalline Intelligenz, die Carroll an zwei Stellen wie folgt definiert: „Crystallized intelligence. This category was used for second-stratum factors with salient loadings on first-stratum factors such as LD (Language development) and V (Verbal ability) that appear to reflect the role of learning and acculturation“ (1993, S. 583). „Crystallized intelligence, concerned with mental processes that reflect not only the operation of fluid intelligence but also the effects of experi-

ence, learning, and acculturation" (1993, S. 624). Neben dieser allgemeinen Definition des Stratum II Faktors Kristalline Intelligenz und des allgemeinen Stratum I Faktors Verbales Verständnis erscheinen zwei weitere Stratum I Faktoren für die Einordnung der Subtests 1 und 2 relevant: „*Lexical knowledge (VL): Extent of vocabulary (nouns, verbs, or adjectives) that can be understood in terms of correct word (semantic) meanings. Although evidence indicates that vocabulary knowledge is a separable component from LD [Language development], it is often difficult to disentangle these two highly correlated abilities in research studies*" (McGrew, 2005, S. 151). Carroll (1993, S. 164) beschreibt Reading decoding (RD) als Stratum I Faktor, da es Beläge gäbe, „that a general skill of word recognition and decoding can be defined factorially independent of some other skills in the language ability domain, and further, that this word recognition skill can be broken down into detailed processes."

Es wird deutlich, dass die Subtests 1 und 2 in diese Klasse von Faktoren passen. Neben den orthographischen Anteilen (V) entspricht insbesondere Subtest 1 dem Lexical Knowledge. In RD werden die für die Erkennung der Anagramme nötigen Prozesse beschrieben: „Grapheme Encoding, Letter Recognition; Multiletter Array Facilitation, Perception of Multiletter Units (e.g. sh, tion); Depth of Processing in Word Recognition; Phonemic Contrast, Decoding" (Carroll, 1993, S. 165).

Dem Stratum II Faktor Fluide Intelligenz wurden Stratum I Faktoren zugeordnet, „whenever the higher salient loadings of first-stratum factors were for factors such as RG (Reasoning) or I (Induction), involving basic intellectual processes of manipulating abstractions, rules, generalizations, and logical relationships" (Carroll, 1993, S. 583). In einer späteren Beschreibung wird noch der geringe Lernanteil an der Ausbildung der entsprechenden Fähigkeiten hervorgehoben: „Fluid Intelligence, concerned with basic processes of reasoning and other mental activities that depend only minimally on learning and acculturation" (S. 624).

Carroll (1993, S. 201ff) führt alle Testarten auf, die er der Fluiden Intelligenz zuordnet. Die für die Einordnung der Subtests 3, 4 und 5 relevanten Beschreibungen lauten wie folgt (die zugehörigen Abkürzungen der Studien wurden im Zitat weggelassen):

Number Series; Number Series Completion [...]: A variety of tasks in which rule-ordered series of numbers are to continued with one or two elements. Items vary in the difficulty of the rule to be discovered. In Number Series II [...] the S is to cross out the number that does not fit in the series.

Form Series [...]: Similar to Number Series, but the stimuli are geometric forms. [...] In Form Series II, S is to cross out the form that does not fit into the series.

Letter Reasoning [...]: Ability to discover a common rule from a series of examples involving series of alphabetical letters.

Letter Series [...]: Find the rule underlying a series of letters, then specify the next two elements. E.g. AR BR CR D?

Letter Sets [...]: Find the one letter group that does not belong in the class of four others; e. g., ABCD LMNO MNOP DEFT UVWX.

Es kann also davon ausgegangen werden, dass die Subtests 3, 4 und 5, die diesen Beschreibungen weitestgehend entsprechen, unter Fluide Intelligenz einzuordnen sind. Carroll unterteilt den Stratum II Faktor Fluide Intelligenz in drei Stratum I Faktoren auf, die je nach verwendetem Test, Stichprobe und Analysemethode unterschiedlich stark interkorrelieren: Sequential Reasoning (RG), Inductive Reasoning (I) und Quantitative Reasoning (RQ). Über alle untersuch-

ten Studien hinweg entsprechen die Buchstabenaufgaben und figuralen Aufgaben in den meisten Fällen dem Induktiven Schlussfolgern (I), während numerische Aufgaben entweder als deduktiv (RG) beurteilt oder (insbesondere bei größerer Schwierigkeit) dem Faktor Quantitatives Schlussfolgern (für das sowohl deduktive als auch induktive Prozesse angenommen werden) zugeteilt werden. Auf eine explizite Zuordnung der Subtests zu diesen drei Kategorien soll aber verzichtet werden.

Die Subtests 6, 7 und 8 können dem Stratum II Faktor Visuelle Wahrnehmung zugeordnet werden. Carroll isoliert bezüglich dieser „broad ability“ einige spezifischere Fähigkeiten. Die den Subtests entsprechenden Faktoren sind (Carroll, 1993):

VZ: Visualization: Ability in manipulating visual patterns, as indicated by level of difficulty and complexity in visual stimulus material that can be handled successfully, without regard to the speed of task solution.

SR: Spatial Relations: Speed in manipulating relatively simple visual patterns, by whatever means (mental rotation, transformation, or otherwise). [...]

P: Perceptual Speed: Speed in finding a known visual pattern, or in accurately comparing one or more patterns, in a visual field such that the patterns are not disguised or obscured. [...]

IM: Imagery: Ability in forming internal mental representations of visual patterns, and in using such representations in solving spatial problems. (S. 362f)

Der Prozess der mentalen Rotation (Subtest 6) wird im Faktor SR (räumliche Beziehungen) explizit erwähnt. Die Aufgabe, die Flächenanzahl dargestellter Körper zu bestimmen (Subtest 7) kann sowohl dem Faktor Visualisation (VZ) als auch dem Faktor Imagination (I) zugeordnet werden. Carroll (1993) bezeichnet es als wünschenswert, dass eine statistische Unterscheidung dieser beiden Konstrukte nachgewiesen wird. Die Beschreibung des Faktors Wahrnehmungsgeschwindigkeit (P) entspricht weitestgehend Subtest 8, so dass erwartet werden kann, dass alle drei Subtests zum Stratum II Faktor Visuelle Wahrnehmung gehören.

Als elementary cognitive tasks (ECT) bezeichnet Carroll (1993, S. 11) Aufgaben, „in which a person undertakes, or assigned, a performance for which there is a specifiable class of ‘successful’ or ‘correct’ outcomes or end states which are to be attained through a relative small number of mental processes or operations“. In den Analysen werden als ECT u. a. folgende Aufgabentypen beschrieben (Carroll, 1993):

Visual search: In this paradigm, a visual stimulus is presented (for example, a particular letter or digit), followed shortly by a series of further stimuli; the subject’s task is to decide whether the initial stimulus is present or absent in the series. Measurements include not only reaction times and accuracies but also the linear regression parameters (intercept, slope) of RTs on the number of stimuli in the series to be searched.

Scan and search: This is a variant of visual search. In Scan and Search, as studied for example by Neisser (1967), the stimulus to be searched for is presented initially there is then a long series of items (each containing a series of stimuli) in which this stimulus is to be searched: the subject’s task is to indicate whether the initial stimulus is present or absent in each set of target stimuli. The task is thus very similar to many that are used in measuring the Perceptual Speed factor, for example the Finding A’s test ... (S. 480)

Die Subtests 9 und 10 entsprechen derartigen Vergleichsprozessen, die Carroll als Processing Speed zusammenfasst.

Subtest 11 wird im Stratum II Faktor „Broad Cognitive Speedness“ als Numerical Facility (N) beschrieben: „Typical Tests of factor N emphasize speed as well as accuracy in handling simple problems. It appears that the most construct-valid tests of N are those that deal with simple problems, that is, addition, subtraction, multiplication, or division of numbers with a small number of digits“ (Carroll, 1993, S. 469). Es kann davon ausgegangen werden, dass die Addition von zehn Ziffern (Subtest 11) dieser Beschreibung entspricht.

Carroll (1993) thematisiert in der Behandlung der Fähigkeiten bezüglich kognitiver Geschwindigkeit das Problem der Differenzierung von Speed und Power (auch Level). Er kommt zu dem Schluss, dass es sich um zwei unterscheidbare Aspekte der Leistungsfähigkeit handelt. Insbesondere für Subtest 9 fällt die Klassifizierung als ECT schwer, da hierbei auch die Sorgfalt eine große Rolle spielt. Unter diesem Aspekt wäre eine Einordnung im Stratum II Faktor Kognitive Geschwindigkeit eher gerechtfertigt. Inwiefern die drei Subtests 9, 10 und 11 auf zwei höheren Faktoren laden, soll jedoch statistisch im Rahmen der Analyse des Modells geklärt werden.

4. Äquivalenz von Papier-Bleistift-Version und Computerversion

4.1 Äquivalenzstudien

Franke (1998) fasste einige Untersuchungen bezüglich der Äquivalenz von Papier-Bleistift-Tests (PBV) mit ihren korrespondierenden Computerversionen (CV) zusammen, um anschließend eigene Ergebnisse von vier Persönlichkeitsinventaren vorzustellen. Für die Persönlichkeitsverfahren SCL (Symptom Checkliste SCL-90-R; Franke, 1995), POMS (Profile of Mood States; Biehl, Dangel & Reiser, 1981), FPI (Freiburger Persönlichkeitsinventar; Fahrenberg, Hampei & Selg, 1994) und IMP (Inventar zur Erfassung interpersonaler Probleme; Horowitz, Strauß & Kordy, 1995) wurde mit Einschränkungen in einzelnen Subskalen prinzipielle Äquivalenz zwischen CV und PBV gefunden. Besonderes Augenmerk soll aber vor allem auf der Zusammenstellung der Äquivalenz-Untersuchungen zu Intelligenz- und Leistungstests liegen (Franke, 1998, S. 47ff). In acht der 24 zitierten Studien werden Matrizentests nach Raven (1958) untersucht. In zwei Studien wurde der Intelligenz-Struktur-Test (Amthauer, 1970) überprüft. Die weiteren Studien betrachten entweder Testbatterien oder im deutschen Sprachraum weniger gängige Verfahren. Bezüglich der Matrizentests fand Franke (1998, S. 52f) folgende Unterschiede bzw. Übereinstimmungen:

Keine Abweichungen der computerunterstützt erhobenen Ergebnisse von 15 untersuchten Kindern zu Handbuchdaten bei RAVEN'S PROGRESSIVE MATRICES (Raven, Court & Court, 1977) fanden Rock und Nolen (1982). Calvert und Waterfall (1982) stellten die Gleichheit beider Versionen von RAVEN'S STANDARD PROGRESSIVE MATRICES (Raven, 1958) bei 83 Schülern fest. Auch Collins und Odell (1986) fanden bei der Untersuchung von nur 16 aphasischen Patienten keine Unterschiede bei RAVEN'S COLOURED PROGRESSIVE MATRICES. 1991 hingegen stellten Neubauer, Urban und Malle bei einer Untersuchung an 100 Studierenden fest, daß die Computeranwendung bei RAVEN'S ADVANCED PROGRESSIVE MATRICES (Kratzmeier, 1976) zu deutlich schlechteren Ergebnissen führte als die herkömmliche Papier-Bleistift-Anwendung. Auch Kubinger und Farkas fanden 1991 schlechtere Testergebnisse unter der Computerbedingung bei 156 Grundwehrdienern in der Anwendung der STANDARD PROGRESSIVE MATRICES. Weibel [1991] allerdings fand bei 66 Schülern keine Unterschiede hinsichtlich der Matrizentests; und auch Norris (1994/1995) belegt die Vergleichbarkeit beider Versionen bei 56 Schülern.

Die Ursachen für solche heterogenen Befunde können in der Stichprobe, in der Umsetzung der Computerversionen, aber auch in der Art der Matrizen (es sind schließlich fünf unterschiedliche Versionen) gesehen werden. Für den Intelligenz-Struktur-Test (Amthauer, 1970) fanden Bulheller (1994) und Weibel (1991) jeweils bessere Bearbeitungsleistungen in der CV. Für die übrigen Verfahren wird in etwa bei der Hälfte Äquivalenz beschrieben; in der anderen Hälfte sind die Probanden entweder in der CV oder der PBV besser. Das Bild ist also verfahrensbedingt uneinheitlich.

Für englischsprachige Verfahren liegen inzwischen einige Meta-Analysen vor. Wang et al. (2007) untersuchten in einer Meta-Analyse von Mathematiktests für Schüler Moderatorvariablen. In einem Regressionsmodell zur Klärung der Mittelwertsunterschiede (operationalisiert als

Effektstärke d) haben die Klassenstufe (Elementar-, Mittel- oder High-School), die Art des Tests (Nationaler Leistungstest, nationaler Fähigkeits- bzw. Fertigkeitstest oder ein Test auf Bundesstaatsebene), die Art der Computerversion (kontrollierte PC-Nutzung oder über das Internet) und die Itempräsentation am PC (linear oder adaptiv) einen bedeutsamen Einfluss. Werden lediglich die korrigierten Mittelwertsunterschiede von PBV und CV verglichen, zeigt sich bei 13 der 44 Studien ebenso ein signifikanter Unterschied wie bei deren Gesamtbetrachtung; bei 18 Studien ist der Betrag des Effekts kleiner gleich 0.05.

Mead und Drasgow (1993) untersuchten metaanalytisch 159 Korrelationen zwischen CV und PBV einzelner Subtests verschiedener Testbatterien aus 28 Studien. Sie unterschieden zwischen „power tests“ und „speeded tests“. Speedtests wurden dabei als Tests mit relativ ähnlichen, prinzipiell von jedem lösbaren Aufgaben definiert. „Timed power tests“ beschreiben Testverfahren, die trotz allgemeiner Zeitgrenze vorwiegend über die Schwierigkeit der Items diskriminieren. Diese wurden jedoch als Powertests deklariert. Die mittlere reliabilitätskorrigierte Korrelation (vgl. Kapitel 8.3) über alle Tests lag bei .90, für die 123 Powertests bei .95 und bei .72 für die 36 Speedtests. Der standardisierte Mittelwertsunterschied über alle Tests betrug -0.04, nur für die Powertests -0.03; das negative Vorzeichen beschreibt, dass die CV etwas schwieriger gewesen zu sein scheinen. Für die Speedtests liegt ein Effekt von 0.15 vor; für diese Art Tests scheint die CV also leichter. In einem Regressionsmodell zur Klärung der Unterschiede erwies sich die Art des Tests (speed oder power) als die bedeutsamste Variable. Des Weiteren zeigte sich die Testbatterie der Studie, aus der der Test entstammt, als relevanter Faktor. Alle weiteren Variablen führten zu keiner bedeutsamen Steigerung der Varianzklärung (maximal .56). Die am ehesten äquivalenten Tests waren die Powertests der ASVAB (Armed Services Vocational Aptitude Battery, $r_{CV-PBV}^* = .97$). Außerdem wurden DAT (Differential Aptitude Tests), MAB (Multidimensional Aptitude Battery; Jackson, 1984) und verschiedene andere Verfahren unterschieden (je $r_{CV-PBV}^* = .90$). In dieser Metaanalyse erwies sich Adaptivität nicht als bedeutsamer Moderator.

Neuman und Baydoun (1998) untersuchten die Äquivalenz von CV und PBV des OST (Office Skills Tests; Science Research Associates, 1984). In ihren Augen sind die zehn Subtests des OST als Speedtests zu klassifizieren, wobei sie aufgrund der Testbeschreibungen auch als „timed power tests“ gelten könnten. Dies ist insofern relevant, da einige Subtests Ähnlichkeiten bezüglich der Subtests des LPS aufweisen. Sie konnten für die Testbatterie eine äquivalente Faktorenstruktur und insgesamt Kongenierität der beiden Modi (also Äquivalenz nach einer linearen Transformation) nachweisen. Die externe Validität bezüglich der von Experten eingeschätzten Jobperformance ist für CV und PBV sehr ähnlich und unterscheidet sich nicht statistisch signifikant. Sowohl die Reliabilität der jeweiligen Modi als auch die Korrelationen untereinander bezüglich der einzelnen Subtests sind als psychometrisch gut bis sehr gut zu werten ($.79 \leq r \leq .98$, $X_{Med} = .90$). Neuman und Baydoun (1998) geben an, dass sich kein Mittelwert auf dem Signifikanzniveau von $\alpha = .0005$ unterscheidet. Die Reanalyse der angegebenen Mittelwerte und Standardabweichungen zeigt jedoch, dass sich sieben der zehn Mittelwerte und drei der zehn Varianzen auf einem Signifikanzniveau von $\alpha = .05$ bedeutsam unterscheiden, wobei in drei Vergleichen die PBV und in vier Vergleichen die CV höhere mittlere Leistungen aufwiesen. Die Wahl des sehr strengen Kriteriums von .0005 wird durch die Kontrolle eines Stichprobeneffekts ($n = 411$) begründet. Da die eigentlich zu bestätigende Hypothese aber die Gleichheit der Mit-

telwerte bzw. Varianzen ist, sollte das Signifikanzniveau eher erhöht anstatt verringert werden (Liebert & Liebert, 1995).

Květon, Jelínek, Vobořil und Klimusová (2007) untersuchten den Einfluss unterschiedlicher Bildschirmdarstellungen und Bearbeitungsreihenfolgen der Modi auf die Äquivalenz. Sie verglichen die PBV der slowenischen Form des Bourdon Tests (Senka, Kuruc & Čečer, 1992) mit zwei CV: graue Symbole auf weißem Grund (regulär) und weiße Symbole auf grauem Grund (invers). Der Bourdon Test ist ein Konzentrationstest ähnlich dem d2 (Brickenkamp, 1972), bei dem unterschiedlich markierte Quadrate zu identifizieren sind. Es zeigen sich teilweise bedeutsame Mittelwertsunterschiede zwischen den beiden CV und der PBV bezüglich fast aller Indizes (Gesamtzahl bearbeiteter Items, Anzahl richtig sowie falsch bearbeiteter Items, relative Performanz, d. h. höchste minus geringste Anzahl bearbeiteter Items einer Zeile dividiert durch die mittlere Anzahl bearbeiteter Items pro Zeile). In der PBV werden weniger Items bearbeitet als in der regulären CV, aber mehr als in der inversen CV. Insgesamt werden in der PBV allerdings am wenigsten Fehler gemacht. Der Reihenfolgeeffekt der Bearbeitungsmodi wurde für den „Test of Concentration of Attention“ (Kučera, 1992) mittels full-crossed Design (d. h. Test und Retest in den Bedingungen PBV-PBV, CV-CV, CV-PBV und PBV-CV) überprüft. Die Aufgabe des Probanden ist es, Zeilen mit zwei mal 15 Zeichen zu vergleichen und unterschiedliche Zeichen zu markieren. Für alle vier Indizes (Gesamtzahl bearbeiteter Zeilen, Anzahl richtig sowie falsch bearbeiteter Zeilen und das Verhältnis von falsch bearbeiteten Zeilen zur Gesamtzahl) zeigen sich Interaktionseffekte bezüglich Reihenfolge und Modus. Die Post-hoc-Analysen zeigen, dass die Leistung in der PBV-PBV-Bedingung konstant bleibt, die Gesamtleistung und die Anzahl richtiger Lösungen in den drei anderen Bedingungen zunehmen sowie die Fehlerzahl nur in der CV-PBV-Bedingung signifikant ansteigt.

4.2 Zur Äquivalenz des LPS

Zum LPS sind nur sehr wenige Äquivalenzuntersuchungen aufzufinden. Eine der ersten Untersuchungen dürfte die Arbeit von Fenske (1992; zitiert nach Hänsgen, 1999) sein, die die beiden Parallelversionen der ersten vier Subtests des LPS an 60 Probanden je einmal in der CV und einmal in der PBV testete. Ohne Berücksichtigung, welche Parallelversion bearbeitet wurde, korrelieren PBV und CV von Subtest 1/2 (Allgemeinbildung) mit .87, von Subtest 3 (Logisches Denken figural) mit .76 und von Subtest 4 (Logisches Denken abstrakt) mit .82. Die Mittelwerte und Standardabweichungen der Modi bei Subtest 1/2 unterscheiden sich signifikant (CV < PBV); für die Subtests 3 und 4 liegen keine bedeutsamen Unterschiede vor. Troche, Rammstedt und Rammsayer (2002) bemerken jedoch, dass in dieser Untersuchung durch die Mittelung über die beiden Parallelformen deren Effekte mit den Effekten der Durchführungsbedingung konfundiert sind. In ihrer Untersuchung bearbeiteten die Probanden entweder zuerst am PC und nach zehnminütiger Pause in der PBV (n = 66) oder umgekehrt (n = 65) jeweils dieselben Aufgaben der vier LPS-Subtests (Parallelform A). Zur Überprüfung der Äquivalenz der beiden Bearbeitungsbedingungen wurde jeweils die erste Testung genutzt, um Lern- und Transfereffekte zu vermeiden. Die Rolle des Stichprobeneffekts bleibt jedoch unerwähnt. Der Mittelwert der PBV von Subtest 1/2 ist um rund sechs Rohwertpunkte höher als in der CV (d = -0.70). Dieser Unterschied ist ebenso signifikant wie bei Subtest 3, wobei hier in der CV im Schnitt zwei Aufgaben

mehr richtig bearbeitet wurden. Für Subtest 4 liegen keine statistisch bedeutsamen Unterschiede vor. Die jeweiligen Modi der Subtests korrelieren bezüglich der Bearbeitungsreihenfolge ähnlich, jedoch nur für Subtest 1/2 in für parallele Versionen angemessener Höhe (Subtest 1/2: $r_{CP} = .79$, $r_{PC} = .87$; Subtest 3: $r_{CP} = .65$, $r_{PC} = .62$; Subtest 4: $r_{CP} = .66$, $r_{PC} = .59$). Die Retestreliabilität der CV ist nur für Subtest 1/2 gut (Subtest 1/2: $r_{CC} = .94$; Subtest 3: $r_{CC} = .74$; Subtest 4: $r_{CC} = .55$).

Eine dritter via PsychINFO auffindbarer Artikel zur Äquivalenz der PBV und CV des LPS ist die Arbeit von Klinck (1998). Die Stichprobe besteht aus für die einzelnen Subtests durchschnittlich 91 beim Arbeitsamt Ratsuchenden (vgl. Tabelle 4.1), denen entweder zuerst die CV und nach einigen weiteren Tests die PBV oder umgekehrt vorgelegt wurde.

Tabelle 4.1: Ergebnisse bezüglich des LPS aus der Studie von Klinck (1998)

Subtest	LPS-neu	n_{PC}	n_{CP}	r_{PC}	r_{CP}	M_P	SD_P	M_C	SD_C	d'
1/2	1	102	139	.86	.89	33.6	10.5	30.2	10.2	0.32*
3	3	106	83	.80	.76	25.9	5.4	23.8	5.6	0.38*
4	4/5	98	83	.64	.76	25.5	3.5	26.9	4.7	-0.34*
7	6	81	87	.68	.65	17.3	6.9	15.0	4.8	0.38*
8	-	88	79	.61	.81	23.0	10.3	28.6	10.8	-0.53*
9	7	89	80	.80	.79	23.4	6.2	25.0	6.6	-0.26
10	8	81	88	.68	.67	23.1	7.7	26.6	6.9	-0.48*
11	-	103	74	.46	.60	20.4	6.0	18.6	4.5	0.33*
12	2	102	73	.82	.65	24.2	6.1	23.4	5.3	0.13

Bemerkung: P = Papier-Bleistift-Version, C = Computerversion; die Reihenfolge der Buchstaben entspricht der Durchführungsreihenfolge.

* markiert signifikante Mittelwertsunterschiede nach den Angaben von Klinck (1998, S. 67, Tabelle 3; $p < .05$). Eine Reanalyse der angegebenen Werte mittels t -Test für unabhängige Stichproben ergibt für alle neun Subtests signifikante Unterschiede ($\alpha = .05$) der Mittelwerte.

Für einige LPS-Subtests (1/2 [1], 3 [3], 7 [6] und 11 [-]; [im LPS-neu]) zeigen sich Vorteile bezüglich der PBV, andere (4 [4/5], 8 [-] und 10 [8]) scheinen in der CV leichter zu sein. Auch hier wurde jeweils die erste Testung verglichen, so dass Stichprobeneffekte prinzipiell nicht ausgeschlossen werden können. Die Korrelationen der beiden Durchführungsbedingungen sind bezüglich der Interpretation als parallele Verfahren teilweise gut und konsistent über die Modi (Subtests 1/2 [1], 3 [3] und 9 [7]), teilweise sind die Zusammenhänge aber zu gering (Subtest 10 [8]) oder inkonsistent (Subtests 8 [-], 11 [-] und 12 [2]; vgl. Tabelle 4.1). Der Unterschied der Standardabweichungen der Subtests 4 [4/5], 7 [6] und 11 [-] ist ebenfalls jeweils statistisch bedeutsam ($\alpha = .05$).

Sowohl Troche et al. (2002) als auch Klinck (1998) diskutieren mögliche Ursachen der Nicht-Äquivalenz der LPS-Subtests. Ein testimmanenter Grund könnte der Speed-Charakter einiger Subtests sein, wie es beispielsweise bei Mead und Drasgow (1993) dargestellt wird. Da die LPS-Subtests jedoch keine Speedtests im eigentlichen Sinne sind, sondern vielmehr timed Powertests wie etwa die OST-Subtests (Neuman & Baydoun, 1998; s. o.), ist es eher unwahrscheinlich, dass diese Begründung als alleinige Erklärung der Effekte dienen kann. Lee, Moreno und Sympton (1986) testeten die Äquivalenz eines arithmetischen Reasoningtests ohne Begrenzung der Bearbeitungsdauer. Hierbei waren die Leistungen in der PBV signifikant ($p < .05$) besser. Troche et al. (2002) vermuten weiterhin, dass die Bindung der Aufmerksamkeit bei der Eingabe der Lösungen, die feste Vorgabe der Bearbeitungsreihenfolge der Items sowie die fehlende Möglich-

keit, frühere Items zu korrigieren, für die schlechteren Leistungen in der CV eine Rolle gespielt haben könnten. Die entzerrte, einzelne Darstellung der Items könnte hingegen als Erleichterung bei der Bearbeitung der CV gesehen werden. Klinck (1998) gibt darüber hinaus an, dass in der CV für einige Subtests die Antwortfelder sowie die Darstellung der Items geändert worden seien und vermutet zu Recht einen Einfluss dieser Änderungen auf die Äquivalenz.

4.3 Bedingungen für Äquivalenz

Bereits 1988 bemängelt Honaker, dass viele Studien zur Überprüfung der Äquivalenz lediglich einen Mittelwertsvergleich der beiden Modi heranziehen. Da aber die Änderung der Itempräsentation auf dem Bildschirm und die Eingabe der Antwort via Tastatur eher zu zwei unterschiedlichen Tests führen als zu äquivalenten Formen eines Tests, sei die gemeinsame Nutzung von Validitätsbelegen und Normdaten „potentially jeopardized“ (S. 561), also gefährdet. Honaker (1988) beschreibt drei Aspekte der Äquivalenz: psychometrische Eigenschaften, experimentelle Äquivalenz und Relativität der Äquivalenz.

4.3.1 Psychometrische Äquivalenz

Zwei Tests sind genau dann äquivalent, wenn sie parallel sind (ITC, 2005; vgl. Kapitel 8.2), wenn sie also gleiche Mittelwerte, gleiche Varianzen, identische Verteilungen und Rangordnungen haben sowie mit anderen Verfahren in derselben Höhe zusammenhängen (Gulliksen, 1950; Lord & Novick, 1968). Unter diesen Bedingungen sind die beiden Verfahren austauschbar; es können sowohl dieselben Normen genutzt als auch dieselben Interpretationen und Schlussfolgerungen gezogen werden. Bei Gültigkeit der Klassischen Testtheorie (vgl. Kapitel 8) sind die Messwerte intervallskaliert und damit bezüglich Mittelwert und Varianz transformierbar. Dies wird beispielsweise in der Angabe von T-Werten ($M = 50$, $SD = 10$), IQ-Werten ($M = 100$, $SD = 15$) oder Prozenträngen deutlich, die allesamt ineinander überführbar sind. Daher ist eine Generalisierung über zwei Tests auch dann möglich, wenn sogenannte essentielle τ -Äquivalenz vorliegt. Dies bedeutet, dass eine akzeptable Korrelation der beiden Tests vorliegt und die Verteilungen der Verfahren lediglich um einen festen Wert voneinander verschoben sind. Eine weitere noch etwas schwächere Annahme stellt die τ -Kongenierität dar. Hierbei ist es zulässig, dass die Messwerte nicht nur um eine Konstante verschoben, sondern auch multiplikativ ineinander überführbar sind. In beiden Fällen ist die Nutzung gemeinsamer Normen eher fraglich, es kann bei Gültigkeit der entsprechenden Annahmen jedoch davon ausgegangen werden, dass von beiden Tests dasselbe Konstrukt gemessen wird (Steyer & Eid, 2001).

Unter der Annahme einer Probabilistischen Testtheorie wäre das Vorgehen bei der Überprüfung der Äquivalenz etwas anders, wobei prinzipiell dieselben Voraussetzungen geprüft werden. Sykes und Ito (1997) zeigen anhand eines Beispiels, welche Parameter zu überprüfen sind. Sie wenden das Raschmodell (Rasch, 1960) an und überprüfen, ob die Rangreihenfolge der Schwierigkeitsparameter der Items (ε) in den beiden Bearbeitungsmodi übereinstimmen. Über eine Regression zur Vorhersage des ε der PBV mittels des ε der CV kontrollieren sie sowohl Zusammenhang (Steigung der Regressionsgleichung) als auch Mittelwertsunterschied (Intercept bzw. Y-Achsenabschnitt). Der Fähigkeitsparameter kann als „differenzenskaliert“ (Irtel, 1996; Rost,

2004) erachtet werden, d. h. additive Transformationen sind zulässig. Somit können die Parameterschätzungen so verschoben werden, dass eine Referenzperson in beiden Modi denselben Fähigkeitswert zugewiesen bekommt. Bei absoluter Äquivalenz würden dann alle Fähigkeitszuweisungen übereinstimmen. Für diese Schätzungen gelten nun wieder die Bedingungen für Äquivalenz der ITC (2005, s. o.).

4.3.2 Experimentelle Äquivalenz

Unter experimenteller Äquivalenz versteht Honaker (1988), dass zwei Verfahren durchaus psychometrisch äquivalent sein können, durch Unterschiede in der Erhebung diese Gleichheit sich statistisch jedoch nicht zeigt. Faktoren, die die experimentelle Äquivalenz beeinflussen, beinhalten emotionale, perzeptuelle und einstellungsbezogene Reaktionen auf die beiden Testformen. Genau diese Aspekte spielen bei der Überprüfung der Äquivalenz von Testung am Computer und Testung mittels Papier und Bleistift eine entscheidende Rolle.

4.3.3 Relativität der Äquivalenz

Honaker (1988) versteht Äquivalenz nicht als absolute Charakteristik eines Tests, sondern unterstellt zudem Populations- und Individuenabhängigkeit. Als Beispiel führt er unterschiedliche Erfahrung mit Computern auf, die ja unabhängig von spezifischen Gruppen vorhanden sein kann. Es ist wichtig zu beachten, dass individuelle aber auch populationsspezifische Unterschiede bei der Untersuchung von Äquivalenz berücksichtigt werden. Die Feststellung der Äquivalenz für eine spezifische Subgruppe der Anwendungspopulation darf nicht unreflektiert auf die Gesamtgruppe übertragen werden.

4.4 Institutionelle Vorgaben

Die International Test Commission (ITC, 2005) hat in den Richtlinien für computerbasierte und über das Internet durchgeführte Tests in Kapitel 2 „Attend to quality issues in CBT and Internet testing“ im Unterpunkt c „Where the CBT/Internet test has been developed from a paper and pencil version, ensure that there is evidence of equivalence“ folgende Forderungen an Testentwickler formuliert (S. 11, übersetzt vom Autor):

1. Es müssen gründlich dokumentierte Nachweise der Äquivalenz von CBT/Internetttest und Papier-Bleistift-Version (wenn die CBT/Internet-Version eine Parallelförmigkeit ist) geliefert werden. Insbesondere ist zu zeigen, dass die beiden Versionen:
 - gleiche Reliabilitäten haben,
 - miteinander in der Reliabilität nach erwarteter Höhe korrelieren,
 - in gleicher Weise mit anderen Tests und externen Kriterien korrelieren und
 - gleiche Mittelwerte und Standardabweichungen aufweisen bzw. durch geeignete Kalibrierung ineinander überführbar sind.
2. Bei der Konstruktion eines CBT/Internettests aus einem Papier-Bleistift-Test ist sicherzustellen, dass:

- der Proband dieselben Kontrollmöglichkeiten hat (z. B. das Auslassen von Items und Ändern von Antworten),
 - die Art der Itempräsentation sicherstellt, dass die Ergebnisse des CBT/Internetttest äquivalent zur PBV sind und
 - das Antwortformat äquivalent ist.
3. Bei Internetttests sollten die Überprüfung der Äquivalenz und die Normierung in derselben Bedingung durchgeführt werden, wie der Test später zur Anwendung kommen soll (z. B. ungesicherte oder unstandardisierte Bedingungen).

Beim Vergleich dieser Forderungen mit den in Kapitel 4.1 zitierten Studien wird deutlich, dass sich diese Kriterien, die in einer ersten, prinzipiell analogen Form von der APA bereits 1986 publiziert wurden, noch nicht durchsetzen konnten. Schoenfeld (1989, S. 20) meinte dazu: „Thus far, little reaction from either test users or test/software publishers has been received...“. Bis dato scheint die Reaktionsbereitschaft nur bedingt größer geworden zu sein, wobei sicherlich einige Ausnahmen aufgeführt werden könnten.

Andererseits zeigt sich, dass das Interesse an der Erforschung der Bedingungen für Äquivalenz nicht besonders groß ist. Linn et al. (2002) formulieren auf Basis der Ergebnisse eines Colloquiums („Computer-Based Testing: Building the Foundation for Future Assessment“, Philadelphia, 1998) eine Forschungsagenda. Abgesehen davon, dass die Forschungsthemen vorwiegend auf die in den USA typischen Leistungstests, die durch Testzentren konstruiert und administriert werden, bezogen sind, finden sich nur zwei Stellen, an denen Fragestellungen, die auf das Thema Äquivalenz bezogen werden können, thematisiert werden (S. 293 f.):

Is it possible to use item statistics from paper-and-pencil administrations to obtain item statistics for CBT administrations? With which types of items will item statistics from paper-and-pencil administrations be the least useful with CBT administrations? [...] How do the test-taking strategies of examinees differ under CBT, CAT, and pencil-and-paper testing conditions? How do these differences affect scores?

Die erste Frage lässt sich leicht bejahen, sobald Äquivalenz der beiden Darbietungsmodi gewährleistet ist; dass dies möglich ist, wurde bereits aufgeführt. Die zweite Frage muss in der deutschen Testtradition inhaltlich umgekehrt werden. Es liegt eine Vielzahl an Tests als PBV vor. *Wie müssen also die Items der CV geschaffen sein, um die Itemparameter der PBV nutzen zu können bzw. damit Äquivalenz vorliegt?* Die dritte Frage wird in dieser Arbeit gestreift werden, ihre Bearbeitung ist jedoch kein vorrangiges Anliegen. Die zweite, neuformulierte Frage stellt die Basis der folgenden Untersuchungen dar.

5. Die Computerversion des neuen LPS

Auf Basis der Überarbeitung des LPS wurde die Computerversion (CV) für möglichst universale Einsetzbarkeit zur Nutzung mittels Browser in den Programmiersprachen html und JavaScript (www.w3.org) sowie php (www.php.net) umgesetzt. Die Items des originalen LPS-Testbogens wurden eingescannt und im Programm Inkscape (Version 0.46, www.inkscape.org) als Vektorgrafik aufgearbeitet, so dass es möglich wurde, die Darstellung der Items bei höchstmöglichem Kontrast ohne Qualitätsverlust jeder Größe anzupassen. Die Umsetzung der CV ermöglicht es, die Testung an jedem PC durchzuführen. Zur Datenerhebung war jedoch zusätzlich eine Internetanbindung notwendig, da die Daten in einer zentralen Datenbank (MySQL, www.mysql.de) gespeichert wurden. In einer Vorstudie wurden zunächst nur Bedingungen untersucht, in denen die Items einzeln präsentiert wurden. Die Hauptuntersuchung wurde um die Darstellung aller Items als Analogie zur Papier-Bleistift-Version (PBV) erweitert.

5.1 Bedingungen der Computerversion

Zur Untersuchung der Effekte verschiedener Darstellungen des Tests am Computer wurde die Präsentation der Items nach drei Bedingungen variiert:

1. Menge: Einzelitempräsentation vs. Gesamtpräsentation
2. Bearbeitbarkeit: Einmaliges Anklicken vs. Bestätigung der Lösung vs. Korrigierbarkeit bis zum Subtestende
3. Zeitvorgabe: Keine Anzeige vs. Anzeige der Gesamtzeit vs. Zeitbeschränkung für jedes Item

Der Aspekt der Darstellung der Items am PC in ihrer Gesamtheit, also analog zur PBV, im Vergleich zur Präsentation jedes Items einzeln ist insbesondere aus zwei Gründen relevant. Zum einen wird in aktuellen Umsetzungen von bestehenden Tests als CV in der Regel die Einzelitemdarstellung gewählt. Zum anderen wäre die Äquivalenz der Einzelitemdarstellung aber im Hinblick auf die Möglichkeit, adaptive Testdurchführung zu nutzen, wünschenswert. Es stellt sich hierbei aber die Frage, ob durch diese offensichtlich deutliche Veränderung des Umfelds des Items eine Beeinflussung der Testleistung zu erwarten ist. Während in der PBV ein Item zumindest von den vorhergehenden und nachfolgenden Items umgeben ist, im LPS sogar von Items der anderen Subtests, wird in der CV in der Einzelitemdarstellung das Item dem Begriff entsprechend alleine abgebildet. Es fehlen somit vermeintlich störende Reize. Andererseits wird der Rückgriff auf bereits bearbeitete Items erschwert. Auf ein bereits angewandtes Lösungsschema kann nur mehr über dessen mentale Repräsentation zurückgegriffen werden. In der PBV ist es hingegen leicht möglich, sich ein bereits bearbeitetes Item wiederholt zu vergegenwärtigen. Außerdem besteht in der PBV die Möglichkeit, ohne großen Zeitverlust Items auszulassen und erst später zu bearbeiten. In der CV vieler Tests gibt es keine Möglichkeit, die Antwort von bereits bearbeiteten Items zu ändern; in einigen Tests ist das Ändern des letzten oder vorletzten Items erlaubt. Aber auch in Tests mit CV, die prinzipiell das Ändern von Antworten zulassen, ist diese Änderung bei Einzelitemdarstellung mit einem größeren Zeitverlust verbunden als in einer PBV. Werden alle Items analog zur PBV auch in der CV gesamt dargestellt, ergibt sich in der Regel das

Problem, dass bei vielen Tests (insbesondere dem LPS) die Zahl der Items so groß ist, dass eine gleichzeitige Darstellung aller Items nicht möglich ist, da die Items in diesem Fall so klein skaliert sein müssten, dass sie nicht oder nur kaum lesbar wären (zumindest auf Monitoren der üblichen Größe). Nur durch Scrollen können dann die ersten oder letzten Items gesehen werden. Auch dies stellt eine vermeintliche Einschränkung im Vergleich zur PBV dar, wobei dieser Effekt als nicht besonders groß erachtet werden kann. Ein Nachteil der Gesamtdarstellung sei noch erwähnt: Im Gegensatz zur Einzelitemdarstellung kann es hierbei passieren, dass Items versehentlich vergessen bzw. ausgelassen werden. Insbesondere bei Tests, deren PBV in ein Testheft und einen Antwortbogen unterteilt ist, kann es zudem passieren, dass durch das Verrutschen auf dem Antwortbogen alle ab diesem Zeitpunkt beantworteten Items als falsch gewertet werden. Bei der Einzelitemdarstellung hingegen ist nur ein bewusstes Auslassen eines Items möglich.

Der zweite Faktor, die Variation der Bearbeitungsmöglichkeiten, ist insofern relevant, als dass hier in verschiedenen Umsetzungen von PBV jeweils Gebrauch gemacht wird. In der Mehrheit der CV, die die Items einzeln darstellen, ist eine Lösung auszuwählen und zu bestätigen, um zum nächsten Item zu gelangen. Einige Verfahren verzichten jedoch auf die Bestätigung. Falls zu bereits bearbeiteten Items zurückgekehrt werden kann, ist dies in der Regel mit der Möglichkeit der Korrektur verbunden. Diese dritte Variante entspricht am ehesten dem Prozedere in der PBV, wo in den meisten Testverfahren Änderungen bereits abgegebener Lösungen zulässig sind. In einer CV mit Einzelitemdarstellung ist eine solche Korrektur jedoch mit einem gewissen Aufwand verbunden, da man ja erst zu dem entsprechenden Item zurückfinden muss. Durch die Verhinderung dieser Möglichkeit wird somit ein derartiger Ressourcenverbrauch unterbunden, was dazu führen kann, dass stattdessen mehrere Items bearbeitet werden und somit zumindest die Wahrscheinlichkeit, weitere richtige Lösungen bzw. Rohpunkte zu sammeln, erhöht wird. Andererseits kann insbesondere die Bedingung, in der ohne Bestätigung zum nächsten Item gewechselt wird, zu Frustration führen, wenn sich der Proband lediglich verlickt und somit seinen Fehler unmittelbar bemerkt. Dies wiederum kann zu einer Erhöhung der Aufmerksamkeit und damit zu mehr korrekten Lösungen führen. Hinsichtlich des Anwendungsaspekts des adaptiven Testens gilt ferner zu bedenken, ob das Zulassen von Korrekturen überhaupt umsetzbar ist. Da die Präsentation von späteren Items von der richtigen bzw. falschen Bearbeitung der vorhergehenden Items abhängt, führt eine Änderung einer bereits abgegebenen Lösung zur Änderung des bereits beschrittenen Entscheidungsbaumes. Durch unmittelbare Berechnung der Personenfähigkeit nach jedem Item sollte diese Situation zwar prinzipiell handhabbar sein, die Änderungen der Parameter der in den zwischen der ersten Lösung und der Korrektur bearbeiteten Items stellen jedoch einen gewissen Aufwand dar, so dass in diesem Sinne Äquivalenz ohne Korrekturmöglichkeit wünschenswerter wäre.

Insbesondere der Aspekt der Aufmerksamkeit und Konzentration wird durch den dritten Faktor berührt: die Vorgabe der Bearbeitungszeit. Alle drei Bedingungen haben Vor- und Nachteile. Keine Zeitangabe entspricht in der Regel der Bearbeitung der PBV. Im Unterschied dazu hat der Proband in der Einzelitemdarstellung (aber auch bedingt in der Gesamtdarstellung) in der CV den Nachteil, dass er einen schlechten Überblick über die bereits bearbeiteten und noch zu bearbeitenden Items hat. Der Vorteil ist, dass die Probanden zum Ende des Tests nicht beginnen können zu raten. Dies stellt sich vornehmlich als Problem der Variante der CV dar, in der die restliche oder abgelaufene Zeit während der Bearbeitung angezeigt wird. Andererseits können die

Items, die nicht unmittelbar zum Schluss bearbeitet werden, ohne den Stress der vermeintlich ablaufenden Bearbeitungszeit gelöst werden. Durch die Vorgabe der Bearbeitungszeit pro Item wird ein gewisser Druck aufgebaut, sobald die Schwierigkeit der Items eine problemlose Bearbeitung ohne Zeitnot nicht mehr zulassen. Demgegenüber forciert eine Begrenzung der Bearbeitungsdauer einzelner Items die Beschäftigung mit einer größeren Anzahl von Items, die dann aber teilweise heuristisch oder durch Raten gelöst werden müssen.

5.2 Ableitung von Hypothesen auf Grundlage der Bedingungen der Computerversion des LPS

Aus dem vorhergehenden Kapitel wird deutlich, dass die variierten Bedingungen der sozusagen unterschiedlichen Computerversionen zumeist sowohl förderliche als auch hinderliche Aspekte bezüglich der Äquivalenz der beiden Modi (CV und PBV) haben können. Es ist zu erwarten, dass die Relevanz des Einflusses zudem abhängig von den speziellen Eigenschaften der jeweiligen Subtests ist. Um eine Flut von Hypothesenformulierungen, die im Extremfall in 165 Paarvergleichen ($3 \times 3 \times 11 + 3 \times 2 \times 11 = 99 + 66 = 165$) bezüglich CV vs. PBV plus den diversen Interaktions- und Haupteffekten resultieren würde, zu vermeiden, wird die Forschungsfrage der Arbeit in allgemeiner Form gestellt, ohne dabei zu prüfende einzelne Hypothesen auszuformulieren.

Hinsichtlich der derzeitigen Praxis der Übertragung von PBV eines Tests auf den Computer wäre der Idealfall, dass unabhängig von der Form der CV, also unter allen Bedingungen, Äquivalenz vorliegt. Im Rahmen der KTT bedeutet Äquivalenz, dass die beiden Testungen (CV und PBV) parallel sind. Zur Nutzung derselben Normen muss strenge Parallelität gefordert werden. Es müssen daher Mittelwert und Varianz überprüft werden und folgenden Zusammenhang erfüllen (vgl. (8-8)):

$$(5-1) \quad M_{PBV} = M_{CV}, Var_{PBV} = Var_{CV}$$

Außerdem muss die Korrelation der Ergebnisse in den Modi ausreichend hoch sein. Nach Lienert und Raatz (1998, S. 269) ist eine angemessene Untergrenze für die Re- und Paralleltest-Reliabilität von standardisierten Leistungstests .80, also:

$$(5-2) \quad r_{PBV-CV} \geq .80$$

Bei der Umsetzung der PBV auf den Computer ist zu erwarten, dass eine fast identische Abbildung des Tests, sprich als Gesamtdarstellung mit Korrekturmöglichkeit und ohne sichtbare Zeitangabe, am ehesten zu äquivalenten Ergebnissen führt, da hierbei lediglich den Computer als solches betreffende Faktoren (z. B. Ängste, Skepsis, Eingabemodalitäten) verzerrend wirken können; diese Einflüsse wirken jedoch in jeder anderen Bedingung der CV auch. Dementsprechend wird in dieser Bedingung am ehesten Äquivalenz erwartet.

6. Inferenzstatistische Methoden

In diesem Kapitel werden die für die Auswertung der Untersuchung genutzten statistischen Verfahren vorgestellt und ihre Auswahl begründet. Ausgespart bleiben die für die Äquivalenzprüfung relevanten Verfahren, die im folgenden Kapitel gesondert beschrieben werden.

6.1 Mittelwertsvergleiche

Neben den Within-Vergleichen der beiden Durchführungsmodi werden auch unterschiedliche Gruppen zu vergleichen sein, z. B. eine Gruppe, die PBV zu T1 durchführt, mit der Gruppe, die die PBV zu T2 durchführt. Üblicherweise werden für Mittelwertsvergleiche unabhängiger Stichproben t -Tests berechnet. Aufgrund der Überlegungen von Kubinger, Rasch und Moder (2009), dass der Welch-Test wegen seiner größeren Robustheit dem t -Test vorzuziehen sei, sollen die möglichen statistischen Verfahren kurz beleuchtet werden. Diehl und Arbinger (2001) stellen für den Mittelwertsvergleich zweier unabhängiger Stichproben den z -Test, den t -Test und den Welch-Test vor. Für sehr kleine Stichproben empfehlen sie den Fisher Randomisationstest. Vom Wilcoxon-Rangsummentest wird insbesondere bei intervallskalierte Daten eher abgeraten.

Für den t -Test gelten die strengsten Voraussetzungen für die zu prüfenden Stichproben. Es wird angenommen, dass die Messgröße in beiden Populationen unabhängig normalverteilt ist. Die Varianz der Messgröße wird als in beiden Grundgesamtheiten identisch (homogen) angesehen. Bei z -Test und W -Test wird die Gleichheit der Varianzen nicht vorausgesetzt. Im z -Test müssen die eigentlich benötigten Varianzen der Grundgesamtheit durch die Stichprobenvarianz geschätzt werden (sofern sie nicht doch bekannt sind). Der W -Test basiert prinzipiell auf der Schätzung durch die Stichprobenvarianzen. Für die valide Wahl des t -Tests müsste also die Normalverteilung der Messgröße in den beiden Stichproben und die Gleichheit der Varianzen überprüft werden. Kubinger et al. (2009) argumentieren nun entsprechend den Ausführungen von Diehl und Arbinger (2001), dass es dadurch zu einer Häufung von Signifikanztests und somit zu einer Kumulierung des α -Fehlers kommt. Da bei Vorliegen der Voraussetzungen für den t -Test bei gleichen Stichprobengrößen dieselben und selbst bei stark abweichenden Stichprobengrößen approximativ die gleichen Ergebnisse für t -Test und W -Test resultieren, sollte auf die Prüfung der Varianzhomogenität verzichtet werden und generell der W -Test angewandt werden. Diehl und Arbinger (2001) führen zudem an, dass der t -Test zwar bei gleichen Stichprobengrößen robust gegenüber dieser Verletzung ist, auf ungleiche Fallzahlen allerdings äußerst empfindlich reagiere. Für große Stichproben ($n_1 > 30$, $n_2 > 30$) empfehlen Diehl und Arbinger (2001) den z -Test, da dessen Prüfgröße standardnormalverteilt ist und somit keine Freiheitsgrade berücksichtigt werden müssen. Zudem nähert sich die Student t -Verteilung für große Fallzahlen ohnehin der Normalverteilung an. Für kleinere Fallzahlen empfehlen auch sie den W -Test. Die Prüfgröße ist für beide Verfahren dieselbe (Diehl & Arbinger, 2001):

$$(6-1) \quad |z| = |W| = \frac{|M_1 - M_2|}{\sqrt{\frac{S_1^2 + S_2^2}{n_1 + n_2}}}$$

Aufgrund der den Voraussetzungen entsprechend unterschiedlichen Annahmen hinsichtlich der Gültigkeit der Schätzung der Populationsvarianz durch die Stichprobenvarianz werden für z und

W unterschiedlich Verteilungen angenommen. z wird als standardisierte Mittelwertsdifferenz zweier normalverteilter Variablen als standardnormalverteilt angesehen, so dass die Wahrscheinlichkeit des resultierenden Wertes unter der $H_0: \mu_1 = \mu_2$ durch die kumulierte Normalverteilung angegeben werden kann ($p = 1 - \Phi(|z|)$). Die Approximation der Verteilung der Prüfgröße W an eine t -Verteilung wird durch die Bestimmung des Freiheitsgrades durch

$$(6-2) \quad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

erreicht. Hiermit kann die Wahrscheinlichkeit, dass die resultierende Differenz von 0 verschieden ist, mittels der t -Verteilung bestimmt werden ($p = 1 - \int_{-\infty}^{|W|} T_{df}(|W|)$, wobei T_{df} der Dichtefunktion der Student t -Verteilung bei df Freiheitsgraden entspricht). Beim Vergleich von zwei Stichproben mit je sechs Versuchspersonen entspricht die Approximation der exakten Verteilung auf drei Nachkommastellen. Mit zunehmender Stichprobengröße wird die Schätzung zunehmend exakter (Welch, 1949).

Diehl und Arbinger (2001) ziehen folgende Schlüsse aus den Studien zur Robustheit von Tests zur Prüfung des Mittelwertsunterschieds zweier unabhängiger Stichproben: Die Abweichung von der Normalverteilung ist unbedenklich, sofern annähernd symmetrische Populationsverteilungen vorliegen. Größere und gleich große Stichproben können die Einbußen an Robustheit aufgrund schiefer Verteilungen teilweise ausgleichen. Bei schiefen Verteilungen sollte das Signifikanzniveau strenger gewählt werden, da das tatsächliche Risiko des α -Fehlers über dem nominellen liegen kann.

6.2 Schiefe und Kurtosis

Soll also auf einen Test auf Normalverteilung verzichtet werden, sollten wenigstens die Schiefe der Stichprobenverteilung betrachtet werden. Es gibt verschiedene Möglichkeiten, die Asymmetrie einer Verteilung zu berechnen. Hopkins und Weeks (1990) empfehlen das standardisierte Moment dritter Ordnung, das durch die Formel

$$(6-3) \quad sk = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$

erwartungstreu und für $n > 150$ approximativ normalverteilt ist mit dem Standardfehler $\sqrt{6/n}$, so dass die Prüfung der $H_0: sk = 0$ möglich wäre. Der Vorteil der Standardisierung ist, dass, da die Größe der Ausprägung maßstabsunabhängig ist, auf die Signifikanztestung und somit die Kumulierung des α -Fehlers verzichtet werden kann; stattdessen wird das Ausmaß der Schiefe anhand der Beurteilungen von Hopkins und Weeks (1990) qualitativ betrachtet. Sie formulieren zwar keine direkte Benennung der Beträge der Schiefe, so wie Cohen (1988) dies hinsichtlich der Bewertung der Effektstärke d in speziellen Fällen vornahm, in einigen Grafiken werden unterschiedliche Werte jedoch klassifiziert. Eine Schiefe von -0.33 wird als „slight negative“ bezeichnet, „moderate“ werden die beiden Werte 0.58 und -0.71 gewertet; 1.15 nennen sie „substantial“, 1.76 „extreme“ und 1.88 „very extreme“. Analog zur Effektstärke könnten die qualitati-

ven Grenzen also bei .3 und .8 gesetzt werden, wodurch Schiefen kleiner .3 als gering, zwischen .3 und .8 als mittel und über .8 als hoch bezeichnet würden⁴.

Ein zweiter Index, der insbesondere die Abweichung von der Normalverteilung beschreibt, ist die Kurtosis (im deutschen Lehrbüchern auch Exzess genannt). Sie wird geschätzt durch die Funktion

$$(6-4) \quad kurt = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{X_i - \bar{X}}{s} \right)^4 - 3 \frac{(n-1)^2}{(n-2)(n-3)},$$

wodurch dieses zentrale Moment vierter Ordnung für große n approximativ normalverteilt ist mit dem Standardfehler $\sqrt{24/n}$ (Hopkins & Weeks, 1990). Gilt $kurt > 0$, wird die Verteilung als leptokurtisch bezeichnet. Im Vergleich zur Standardnormalverteilung treten die äußeren Bereiche häufiger auf. Zur Wahrung der Varianz (die ja auf 1 normiert wurde) folgt in der Regel ein höherer Peak der Verteilung (DeCarlo, 1997). Eine Verteilung mit weniger extremen Werten als die Normalverteilung heißt platykurtisch. Demensprechend haben platykurtische Verteilungen, insbesondere bei Vorliegen der Symmetrie, im mittleren und im extremen Bereich meist eine geringere Dichte als die Normalverteilung. Schiefe Verteilungen sind in aller Regel leptokurtisch.

Für den Vergleich von Mittelwerten bestehen die Konsequenzen bei Vorliegen von Schiefe und Kurtosis in der beeinflussten Robustheit des Tests durch eine Verschiebung des tatsächlichen α -Fehlers. Liegt eine platykurtische Verteilung ($kurt < 0$) vor, folgt daraus, dass der das α -Niveau definierende kritische Wert einen höheren p -Value aufweist als α . Die Wahrscheinlichkeit, einen Fehler erster Art zu machen, erhöht sich also. Der Einfluss einer positiven Kurtosis hängt beim Mittelwertsvergleich von der Symmetrie der zu vergleichenden Stichproben ab. Positive Schiefewerte werden als linkssteil bezeichnet, negative als rechtssteil. Durch die dritte Potenz kommt der Differenz zum Mittelwert von extremen Werten ein stärkeres Gewicht zu, weshalb bei einer rechtssteilen Verteilung die nach der Standardisierung negativen linken Werte den entsprechenden Ausschlag geben (Bortz, 2005).

⁴ Entsprechende dem vom Hopkins und Weeks (1990) formulierten Signifikanztest ist .8 der kritische Wert, an dem die $H_0: sk = 0$ mit dem α -Risiko von 5 % abgelehnt werden muss, wenn $n = 25$; der Wert .3 bildet bei $n = 181$ die Grenze für das α -Niveau von 5 %.

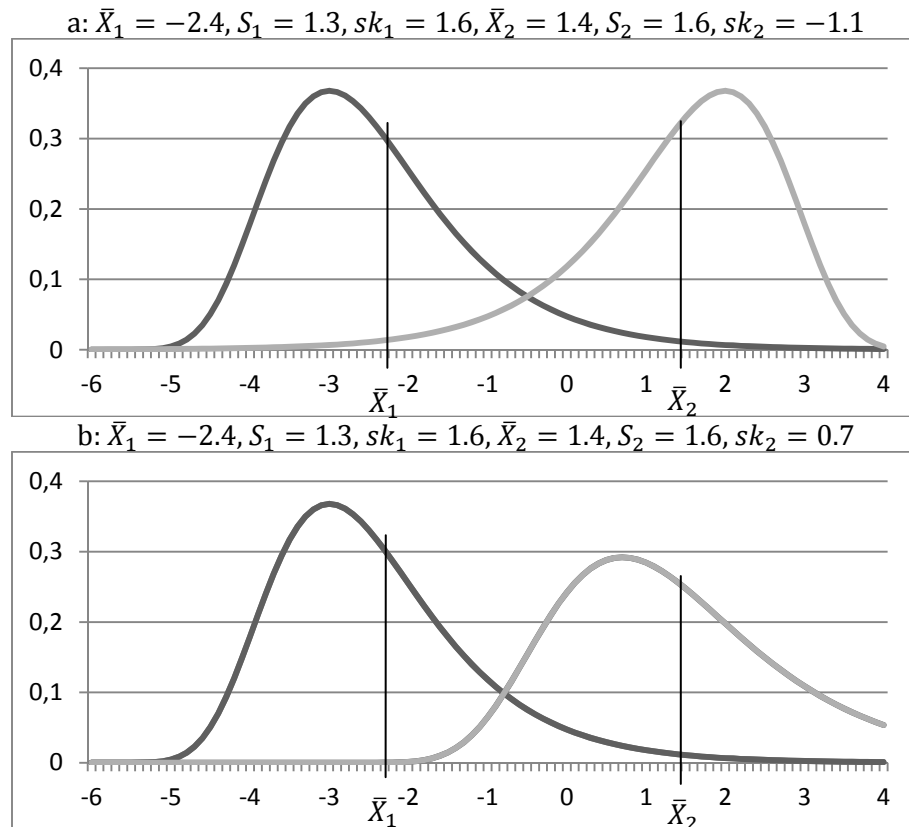


Abbildung 6.1: Schiefe Verteilungen (Fisher-Tippett-Verteilungen) zur Illustration des Einflusses der Schiefe

Abbildung 6.1 veranschaulicht den Einfluss der Schiefe einer Verteilung beim Vergleich der Mittelwerte. Der obere rechte Graph einer rechtssteilen Verteilung ($sk_2 = -1.1$) zeigt aufgrund seiner negativen Schiefe bei gleichem Mittelwert und gleicher Varianz einen deutlich geringeren Überschneidungsbereich als der untere. Ist die Verteilung der Stichprobe mit dem deskriptiv größeren Mittelwert also rechtssteil (a), liegt das tatsächliche α -Risiko über dem nominellen. Werden Aussagen also auf dem 5 % Signifikanzniveau getroffen, ist das Risiko, die Nullhypothese fälschlich abzulehnen, tatsächlich höher. Analog ist der wahre p -Value höher, als er vom Statistikprogramm angegeben wird. Umgekehrt können bei einer linkssteilen Verteilung (b) Aussagen auf dem 5 % Signifikanzniveau als konservativer betrachtet werden. Hopkins und Weeks (1990) schlagen für (a) eine Verringerung der α -Grenze vor. Ist die Verteilung der ersten Stichprobe bzw. die Stichprobe mit dem kleineren deskriptiven Mittelwert rechtssteil, hat dies immer eine Erhöhung des tatsächlichen α -Risikos zur Folge.

7. Verfahren zur Überprüfung der Äquivalenz

Zwei Verfahren sind dann äquivalent, wenn sie im Sinne der Klassischen Testtheorie parallel sind (ITC, 2005). In der vorliegenden Untersuchung werden die elf Subtests des LPS-neu bezüglich ihrer Äquivalenz in Papier-Bleistift-Version (PBV) und Computerversion (CV) überprüft. Mittels der CV werden unterschiedliche Bedingungen zur Überprüfung deren Einflusses realisiert. Variiert werden neben dem Durchführungsmodus die Bearbeitungsreihenfolge (entweder erst CV, dann PBV oder erst PBV, dann CV), die Präsentationsart (Einzelitemdarstellung oder Gesamtdarstellung), die Bearbeitungsmodalität (nicht, einmalig oder bis Testende korrigierbar) und die Visualisierung der Zeitvorgabe (ohne, gesamt oder Zeitbegrenzung pro Item). Diese fünf unabhängigen Bedingungen werden bei jedem Probanden je Subtest variiert, so dass jeder Subtest als abhängige Variable betrachtet werden muss. Die Variation der Bedingungen der CV über die Subtests hinweg hat einen Vor- und einen Nachteil. Der Nachteil besteht darin, dass es somit nicht mehr möglich ist, die gängigen multivariaten Verfahren anzuwenden, da hierfür für jede abhängige Variable dieselben Faktorausprägungen vorausgesetzt werden. Der Vorteil ist allerdings darin zu sehen, dass durch die Variation die Konfundierung von Stichprobe und Effekten auf ein Minimum reduziert wird. Da jede Versuchsperson elf der 15 Faktorkombinationen randomisiert ausgesetzt wird, können die Effekte der unabhängigen Variablen quasi anhand von elf Messungen untersucht werden.

7.1 Unterschiedlichkeit der Bedingungen der CV

Bevor die Äquivalenz von CV und PBV betrachtet wird, soll zuerst die Unterschiedlichkeit der Bedingungen der CV untersucht werden. Zunächst werden univariate Varianzanalysen mit den vier Faktoren Bearbeitungsreihenfolge, Präsentationsart, Bearbeitungsmodalität und Zeitvorgabe für jeden Subtest berechnet. Für die Interpretation der Effekte soll ein Effektmaß dienen. Da es nicht praktikabel erscheint, über eine Vielzahl von Paarvergleichen die klassischen Maße der standardisierten Mittelwertsunterschiede (Cohen, 1969, 1988; Glass, 1976; Hedges, 1981) zu be- und schließlich zu verrechnen, wird stattdessen der Anteil der erklärten Varianz je Effekt betrachtet. Die drei üblichen Maße hierfür sind η^2 (Pearson, 1905, zit. nach Olejnik & Algina, 2000), ε^2 (Kelley, 1935) und ω^2 (Hays, 1963); ε^2 und ω^2 dienen zur Schätzung des Populations-effektes, wobei ω^2 eine Adjustierung von ε^2 darstellt. η^2 beschreibt den Effekt innerhalb einer Stichprobe (Olejnik & Algina, 2000). Da in dieser Arbeit die Effekte der Bedingungen hinsichtlich der elf Subtests des LPS-neu diskutiert werden, ist eine Adjustierung bzgl. der Population vorerst unnötig, so dass die Verrechnungen für eine erwartungstreue Schätzung zugunsten der direkten Interpretation des Anteils erklärter Varianz unterbleiben kann. Pierce, Block und Aguinis (2004) weisen in ihrer Arbeit darauf hin, dass in vielen Studien das partielle η^2 mit dem klassischen η^2 verwechselt wird und es so zu einer Überschätzung der tatsächlichen Effekte kommt. Cohen (1973), der vermutlich die Begriffe „klassisch“ und insbesondere „partiell“ begründet hat, beschreibt, für welchen Zweck welches Effektmaß geeignet ist. Werden Variablen systematisch variiert, kann ihr Einfluss nicht als Fehler gesehen werden und findet somit keinen Eingang in die Residual- (oder Fehler-) Varianz. Daher sollte in einem Setting für die Effekte einzelner Fak-

toren und Faktorkombinationen das partielle η^2 genutzt werden, da dann der Einfluss der übrigen Faktoren entfernt („auspartialisiert“) und so ein Vergleich des Effektes über verschiedene Experimente hinweg ermöglicht wird. Auch der Einfluss konstant gehaltener Variablen soll auspartialisiert werden, da auch sie die Residualvarianz nicht beeinflussen. Werden hingegen Variablen als Kontrollvariablen eingeführt, gehören sie ja per Definition zur (allerdings kontrollierten) Residualvarianz im experimentellen Setting. In diesem Fall ist das klassische η^2 zu verwenden. In gemischten Designs mit experimentellen Faktoren und Kontrollvariablen sollte dementsprechend ein dem Design angepasstes („custom-tailored partial η^2 “, Cohen, 1983, S. 111) Effektmaß genutzt werden. η_p^2 ist das Verhältnis der Quadratsumme des Faktors zur Summe der Quadratsummen dieses Faktors und der Residual-Quadratsumme (z. B. Cohen, 1973):

$$(7-1) \quad \eta_p^2 = \frac{QS_A}{QS_A + QS_R}.$$

Dieser Logik folgend werden zum Vergleich der Effekte der Durchführungsbedingung der CV die partiellen η^2 berechnet. Dass die Schätzung von η^2 zur Überschätzung tendiert, wobei der Fehler für große Stichproben geringer wird (Grissom & Kim, 2005), führt im Kontext des Ziels, möglichst die Gleichheit zu belegen, zu unkritischen Konsequenzen, da die Schlussfolgerungen somit konservativer werden.

Tabelle 7.1: Designmatrix

Modus	PBV	CV														
Präsentation		Einzel									Alle					
Antwortformat		Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar	
Zeitvorgabe		Ohne	Gesamt	Item	Ohne	Gesamt	Item	Ohne	Gesamt	Item	Ohne	Gesamt	Ohne	Gesamt	Ohne	Gesamt
Bedingung		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Reihenfolge	T1															
	T2															

Bemerkungen: PBV: Papier-Bleistift-Version; CV: Computer-Version; Einzel: Jedes Item wird für sich dargestellt und bearbeitet; Alle: Alle Items des Subtests werden analog zur PBV dargestellt; Einmalig: Mit der Auswahl der Lösung erscheint das nächste Item; Weiter: Erst nach der Bestätigung der Eingabe erscheint das nächste Item; Korrigierbar: Über ein Menü ist während der gesamten Testbearbeitung jedes Item aufrufbar; Ohne: Die Vp sieht keine Zeitvorgabe; Gesamt: Die Vp sieht die verbleibende Bearbeitungszeit; Item: Die Vp sieht eine begrenzte Bearbeitungszeit pro Item.

Obwohl die Darstellung eine Verschachtelung von Antwortformat und Zeitvorgabe impliziert, gilt diese nur auf der Ebene der Präsentation, da bei der Gesamtdarstellung keine Itembearbeitungszeit umgesetzt wurde.

Der Übersichtlichkeit halber wurde darauf verzichtet, alle Kombinationen bezüglich der Reihenfolge des Testmodus aufzuschlüsseln. Unabhängig davon, ob die SV als erste oder zweite Testung durchgeführt wurde, wurden alle darunter stehenden Bedingungen in gleicher Weise variiert.

In der Konsequenz ergeben sich im ersten Schritt auf diese Weise pro Subtest 15 Effekte (vier Haupteffekte, sechs 2-fach, vier 3-fach und eine 4-fach Interaktion, vgl. Tabelle 7.1). Für die einzelnen Effekte bzw. die η_p^2 der Bedingungen bzw. Bedingungskonstellationen ist es jedoch möglich, Konfidenzintervalle anzugeben. Hierfür wird die Theorie der auf Nichtzentralität basierenden Intervallschätzung den Ausführungen von Steiger und Fouladi (1997) und Steiger (2004) folgend eingeführt.

Die F -Verteilung ist ein Spezialfall der nichtzentralen F -Verteilung, in dem der Zentralitätsparameter gleich Null ist. Die nichtzentrale F -Verteilung hat neben den beiden Freiheitsgraden von Zähler und Nenner einen weiteren Parameter, den Zentralitätsparameter λ . Das

Besondere an λ ist, dass unter Gültigkeit der H_0 λ gleich Null ist; je unwahrscheinlicher der in H_0 postulierte Zusammenhang ist, desto größer wird λ , solange das α -Niveau und die Stichprobengröße gleich bleiben. Für die Bestimmung eines Konfidenzintervalls für η_p^2 kann nun dadurch das Problem der Abhängigkeit der Verteilung der Stichprobenschätzungen von dem Wert der Populationsgröße (Fisher, 1928; Wishart, 1932) umgangen werden. Dies ermöglicht das Konfidenzintervall-Transformationsprinzip und das Konfidenzintervall-Inversionsprinzip (Steiger, 2004, S. 167f):

Konfidenzintervall-Transformationsprinzip (confidence interval transformation principle): Sei $f(\theta)$ eine streng monotone Funktion von θ , l_1 und l_2 untere und obere Grenze eines $1 - \alpha$ Konfidenzintervalls von θ . Wenn nun $f(\theta)$ streng monoton steigend ist, sind $f(l_1)$ und $f(l_2)$ untere und obere Grenze eines $1 - \alpha$ Konfidenzintervalls von $f(\theta)$. Wenn dagegen $f(\theta)$ streng monoton fallend ist, sind $f(l_2)$ und $f(l_1)$ untere und obere Grenze eines $1 - \alpha$ Konfidenzintervalls von $f(\theta)$.

Als Beispiel führt Steiger (2004) das Konfidenzintervall für eine Varianzschätzung an. Falls nun anstatt der Varianz die Standardabweichung angegeben werden soll, genügt es somit, die Wurzel der Intervallgrenzen zu ziehen.

Konfidenzintervall-Inversionsprinzip (inversion confidence interval principle): Sei x die Realisation der Zufallsvariable X mit der kumulierten Verteilungsfunktion $F(x, \lambda) = \Pr(X \leq x | \lambda)$ für ein beliebiges numerisches λ . Seien $\alpha_1 + \alpha_2 = \alpha$ mit $0 < \alpha < 1$ feste Werte. Wenn $F(x, \lambda)$ streng monoton fallend ist für λ bei festem x , wähle $l_1(x)$ und $l_2(x)$ so, dass $\Pr(X \leq x | \lambda = l_1(x)) = 1 - \alpha_2$ und $\Pr(X \leq x | \lambda = l_2(x)) = \alpha_1$. Wenn $F(x, \lambda)$ streng monoton steigend ist für λ bei festem x , wähle $l_1(x)$ und $l_2(x)$ so, dass $\Pr(X \leq x | \lambda = l_1(x)) = \alpha_1$ und $\Pr(X \leq x | \lambda = l_2(x)) = 1 - \alpha_2$. Dann ist das Intervall $[l_1(x); l_2(x)]$ ein $1 - \alpha$ Konfidenzintervall von λ . Der Beweis für diesen Satz wird von Casella und Berger (2008, S. 430ff) unter dem Titel „Pivoting the CDF“ (also in etwa „Ausbalancieren der kumulierten Dichteverteilung“) geführt. Da die nichtzentrale F -Verteilung für konstante Freiheitsgrade streng monoton fallend in λ ist, kann damit ein Konfidenzintervall für λ bestimmt werden. Da der Zusammenhang

$$(7-2) \quad \eta_{p(A)}^2 = \frac{\lambda_A}{\lambda_A + N} \text{ bzw. } \lambda_A = \frac{N\eta_{p(A)}^2}{1 - \eta_{p(A)}^2}$$

für einen Faktor A gilt (Fleishman, 1980), wobei N für die Gesamtzahl der Beobachtungen (in der Regel die Stichprobengröße) steht, kann aufgrund des Konfidenzintervall-Transformationsprinzips ein Konfidenzintervall für η_p^2 bestimmt werden. Wie üblich wird $\alpha_1 = \alpha_2 = \alpha/2$ gesetzt, so dass die Irrtumswahrscheinlichkeit um das Intervall herum symmetrisch ist. Das Intervall selbst hingegen wird umso unsymmetrischer, je unwahrscheinlicher die H_0 wird. Basierend auf seinem Buch stellt Smithson (2003) Syntax-Skripte zur Berechnung der Intervallgrenzen von η^2 bzw. η_p^2 mittels SPSS zur Verfügung, die für diese Arbeit genutzt werden⁵.

Es konnten in der Literatur keine Methoden zur gleichzeitigen Erfassung mehrerer (Varianz-) Effekte verschiedener Variablen einer Stichprobe – und damit abhängige Effekte – gefunden werden. Weder in den klassischen Lehrbüchern zu Meta-Analysen (Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Rosenthal, 1984), noch in neueren Auflagen und Zusammenstellungen

⁵ <http://psychology3.anu.edu.au/people/smithson/details/CIstuff/CI.html> [Stand: 02.07.2010]

(Raudenbush & Bryk, 2006; Rustenbach, 2003; Schulze, 2004) werden entsprechende Methoden abgehandelt. Der Mittelwert der Effekte über alle elf Subtests belegt zumindest, inwieweit der generelle Einfluss einer Bedingung bzw. Bedingungskonstellation überhaupt eine Rolle spielen könnte. Der zentrale Grenzwertsatz der Statistik (z. B. Olkin, Gleser & Derman, 1980) besagt, dass die Summe von unabhängigen Zufallsvariablen (und damit auch der Mittelwert) unabhängig von ihrer ursprünglichen Verteilung approximativ normalverteilt ist. Diesbezüglich müssen jedoch zwei Einschränkungen in Kauf genommen werden. Zum einen kann nur in den Bedingungskonstellationen der 3-fach und der 4-fach Interaktion von unabhängigen Zufallsvariablen ausgegangen werden, zum anderen sind die elf Ziehungen resp. Subtests eine etwas geringe Anzahl für die Approximation. Dennoch wird der Bemerkung von Conover (1999, S. 58), dass „sometimes when n is as small as 5 or 10 the normal approximation can be quite good“, folgend und in Ermangelung einer Alternative von diesem Satz Gebrauch gemacht und darauf basierend ein Konfidenzintervall der gemittelten η_p^2 pro Effekt bestimmt, also:

$$(7-3) \quad (1 - \alpha) - \text{KI von } M(\eta_{p(\text{Effekt})}^2) = M(\check{\eta}_{p(\text{Effekt})}^2) \pm \frac{S_{\check{\eta}_{p(\text{Effekt})}^2} \cdot z(1-\alpha/2)}{\sqrt{n}},$$

mit dessen Hilfe die Bedeutsamkeit der Bedingungsvariation abgeschätzt werden kann.

7.2 Unterschiedlichkeit von PBV und CV

Da für den Faktor „Modus“ außer der Testungsreihenfolge nur in der CV die übrigen Faktoren variiert werden, spielen in der Varianzanalyse hinsichtlich der Äquivalenz nur die Haupteffekte Modus und Reihenfolge eine Rolle. Die bezüglich der CV relevanten Effekte betreffen stattdessen immer eine Interaktion mit den Faktoren Modus und Testreihenfolge (vgl. Tabelle 7.2).

Zur Überprüfung der Frage, ob die Itempräsentation unabhängig von Antwortformat und Zeitvorgaben einen Effekt bezüglich der Äquivalenz hat, muss beispielsweise die Interaktion Modus \times Reihenfolge \times Präsentation bezüglich der Subtests betrachtet und gegebenenfalls hinsichtlich der Reihenfolge aufgeschlüsselt werden. Insgesamt wird es häufig notwendig sein, für einzelne Aspekte gesonderte Berechnungen anzustellen, da insbesondere auf Subtestebene auch die univariate Betrachtung von Interesse ist.

Tabelle 7.2: Aufteilung der Stichprobengruppe (G1 bis G15) hinsichtlich der Durchführung der Subtests in den jeweiligen Bedingungen der CV

Modus		PBV	CV														
Präsentation			Einzel									Alle					
Antwortformat			Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar	
Zeitvorgabe			Ohne	Gesamt	Item	Ohne	Gesamt	Item	Ohne	Gesamt	Item	Ohne	Gesamt	Ohne	Gesamt	Ohne	Gesamt
Bedingung			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Subtest	1	A	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15
	2	A	G15	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14
	3	A	G14	G15	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13
	4	A	G13	G14	G15	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12
	5	A	G12	G13	G14	G15	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
	6	A	G11	G12	G13	G14	G15	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10
	7	A	G10	G11	G12	G13	G14	G15	G1	G2	G3	G4	G5	G6	G7	G8	G9
	8	A	G9	G10	G11	G12	G13	G14	G15	G1	G2	G3	G4	G5	G6	G7	G8
	9	A	G4	G8	-	G11	G9,G12	-	G10,G13	G14	-	-	-	G1,G15	G2,G5	G3,G6	G7
	10	A	G7	G4	-	G6	G8,G11	-	G9,G12	G10,G13	-	-	-	G14	G1,G15	G2	G3,G5
	11	A	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G1

Bemerkungen: A symbolisiert die Durchführung der PBV durch alle Vpn. Zusätzlich muss noch innerhalb der Gruppen zwischen denjenigen, die die PBV, und denjenigen, die die CV zuerst bearbeitet haben, unterschieden werden. Dieser Faktor wird zugunsten der Übersichtlichkeit nicht abgebildet.

Nicht signifikante Effekte stellen in ihrer Konsequenz (kein Mittelwertsunterschied) einen ersten Hinweis auf Äquivalenz dar. Da die Durchführungsbedingungen für die einzelnen Subtests des LPS-neu für eine Versuchsperson unterschiedlich waren, können die Ergebnisse nicht mit der klassischen multivariaten Methodik untersucht werden. Stattdessen werden die standardisierten Mittelwertsunterschiede hinsichtlich der PBV und der CV nach einem multivariaten Modell von Hedges und Olkin (1985, S. 206ff) betrachtet. Das Modell soll im Folgenden stark vereinfacht und auf das gegebene Setting angepasst vorgestellt werden:

Für die jeweiligen Durchführungsbedingungen wird basierend auf Tabelle 7.3 und Tabelle 7.4 der Effektivvektor wie folgt berechnet:

$$(7-4) \quad \mathbf{d}_i = (d_{i1}, \dots, d_{ip}, \dots, d_{i11})'$$

mit

$$(7-5) \quad d_{il} = c \frac{\bar{Y}_{il}^{CV} - \bar{Y}_{il}^{PBV}}{s_{il}^{(i)}}, \text{ mit } c = 1 - \frac{3}{4(n_i^{CV} + n_i^{PBV} - 2) - 1},$$

$$i = 1, \dots, k, \dots, 15, l = 1, \dots, p, \dots, 11,$$

wobei \bar{Y}_{il}^{CV} und \bar{Y}_{il}^{PBV} die Mittelwerte der CV und der PBV von Subtest l in Bedingung i , $s_{il}^{(i)}$ das l -te Diagonalelement der gepoolten Kovarianzmatrix $\mathbf{S}^{(i)}$ (also die Varianz) und c ein Korrekturfaktor für kleine Stichprobengrößen, der für größere Stichproben gegen 1 strebt, sind.

Tabelle 7.3: Datenlayout (nach Hedges & Olkin, 1985, S. 207, angepasst an die vorliegende Untersuchung)

Bedingung	CV						PBV					
	Vp	Subtest					Vp	Subtest				
		1	...	p	...	11		1	...	p	...	11
1	1	Y_{111}^{CV}	...	Y_{1p1}^{CV}	...	Y_{111}^{CV}	1	Y_{111}^{PBV}	...	Y_{1p1}^{PBV}	...	Y_{111}^{PBV}
⋮	⋮	⋮		⋮		⋮	⋮	⋮		⋮		⋮
1	n_1^{CV}	$Y_{11n_1^{CV}}^{CV}$...	$Y_{1pn_1^{CV}}^{CV}$...	$Y_{111n_1^{CV}}^{CV}$	n_1^{PBV}	$Y_{11n_1^{PBV}}^{PBV}$...	$Y_{1pn_1^{PBV}}^{PBV}$...	$Y_{111n_1^{PBV}}^{PBV}$
⋮	⋮	⋮		⋮		⋮	⋮	⋮		⋮		⋮
k	1	Y_{k11}^{CV}	...	Y_{kp1}^{CV}	...	Y_{k11}^{CV}	1	Y_{k11}^{PBV}	...	Y_{kp1}^{PBV}	...	Y_{k11}^{PBV}
⋮	⋮	⋮		⋮		⋮	⋮	⋮		⋮		⋮
k	n_k^{CV}	$Y_{k1n_k^{CV}}^{CV}$...	$Y_{kpn_k^{CV}}^{CV}$...	$Y_{k11n_k^{CV}}^{CV}$	n_k^{PBV}	$Y_{k1n_k^{PBV}}^{PBV}$...	$Y_{kpn_k^{PBV}}^{PBV}$...	$Y_{k11n_k^{PBV}}^{PBV}$
⋮	⋮	⋮		⋮		⋮	⋮	⋮		⋮		⋮
15	1	Y_{1511}^{CV}	...	Y_{15p1}^{CV}	...	Y_{1511}^{CV}	1	Y_{1511}^{PBV}	...	Y_{15p1}^{PBV}	...	Y_{1511}^{PBV}
⋮	⋮	⋮		⋮		⋮	⋮	⋮		⋮		⋮
15	n_{15}^{CV}	$Y_{151n_{15}^{CV}}^{CV}$...	$Y_{15pn_{15}^{CV}}^{CV}$...	$Y_{1511n_{15}^{CV}}^{CV}$	n_{15}^{PBV}	$Y_{151n_{15}^{PBV}}^{PBV}$...	$Y_{15pn_{15}^{PBV}}^{PBV}$...	$Y_{1511n_{15}^{PBV}}^{PBV}$

Tabelle 7.4: Parameterschätzer (nach Hedges & Olkin, 1985, S. 208, angepasst an die vorliegende Untersuchung)

Bedingung	Mittelwerte		Kovarianzmatrix	Korrelationsmatrix
	CV	PBV		
1	$\bar{Y}_1^{CV} = (\bar{Y}_{11}^{CV}, \dots, \bar{Y}_{1p}^{CV}, \dots, \bar{Y}_{111}^{CV})'$	$\bar{Y}_1^{PBV} = (\bar{Y}_{11}^{PBV}, \dots, \bar{Y}_{1p}^{PBV}, \dots, \bar{Y}_{111}^{PBV})'$	$\mathbf{S}^{(1)}$	$\mathbf{R}^{(1)}$
⋮	⋮	⋮		
k	$\bar{Y}_k^{CV} = (\bar{Y}_{k1}^{CV}, \dots, \bar{Y}_{kp}^{CV}, \dots, \bar{Y}_{k11}^{CV})'$	$\bar{Y}_k^{PBV} = (\bar{Y}_{k1}^{PBV}, \dots, \bar{Y}_{kp}^{PBV}, \dots, \bar{Y}_{k11}^{PBV})'$	$\mathbf{S}^{(k)}$	$\mathbf{R}^{(k)}$
⋮	⋮	⋮		
15	$\bar{Y}_{15}^{CV} = (\bar{Y}_{151}^{CV}, \dots, \bar{Y}_{15p}^{CV}, \dots, \bar{Y}_{1511}^{CV})'$	$\bar{Y}_{15}^{PBV} = (\bar{Y}_{151}^{PBV}, \dots, \bar{Y}_{15p}^{PBV}, \dots, \bar{Y}_{1511}^{PBV})'$	$\mathbf{S}^{(15)}$	$\mathbf{R}^{(15)}$

Bemerkungen: $\mathbf{S}^{(k)}$ ist die gepoolte Kovarianzmatrix der elf Subtests bezüglich der Bedingung k aus den beiden Kovarianzmatrizen der CV und der PBV. Die mittlere Streuung wird durch $s_{pp}^{(k)} = \sqrt{\frac{(n^{CV}-1)(s^{CV})^2 + (n^{PBV}-1)(s^{PBV})^2}{n^{CV} + n^{PBV} - 2}}$ berechnet.

Als nächstes muss geprüft werden, ob die Effektgrößen einer Bedingung für alle Subtests gleich sind, also ob

$$(7-6) \quad \delta_{i1} = \dots = \delta_{ip} = \dots = \delta_{i11} = \delta_i^*$$

gilt. Hierfür kann die mit $p - 1$ Freiheitsgrade approximativ χ^2 -verteilte Prüfgröße

$$(7-7) \quad Q_i = \mathbf{d}_i' \mathbf{M}_i \mathbf{d}_i, \text{ mit } \mathbf{M}_i = \mathbf{\Lambda}^{(i)} - \frac{\mathbf{\Lambda}^{(i)} \mathbf{e} \mathbf{e}' \mathbf{\Lambda}^{(i)}}{\mathbf{e}' \mathbf{\Lambda}^{(i)} \mathbf{e}}$$

berechnet werden. \mathbf{e} ist ein p -(hier also 11-)dimensionaler Einheitsvektor; $\mathbf{\Lambda}^{(i)}$ ist die Inverse der Kovarianzmatrix $\mathbf{\Sigma}^{(i)} = (\sigma_{st}^{(i)})$, deren Elemente durch

$$(7-8) \quad \sigma_{ss}^{(i)} = \frac{n_i^{CV} + n_i^{PBV}}{n_i^{CV} n_i^{PBV}} + \frac{\delta_{is}^2}{2(n_i^{CV} + n_i^{PBV})} \text{ und}$$

$$(7-9) \quad \sigma_{st}^{(i)} = \rho_{st}^{(i)} \sqrt{\sigma_{ss}^{(i)} \sigma_{tt}^{(i)}},$$

berechnet werden, wobei $\rho_{st}^{(i)}$ für die Korrelation zwischen Subtest s und Subtest t in Bedingung i (durch $r_{st}^{(i)}$ geschätzt), $\sigma_{ss}^{(i)}$ für die Diagonale der Matrix $\Sigma^{(i)}$, also die Varianzen der Subtests und $\sigma_{st}^{(i)}$ für die Kovarianzen der Subtests s und t stehen. Die Kovarianzmatrix $\Sigma^{(i)}$ kann durch das Produkt der Korrelationsmatrix mit der Diagonalmatrix \mathbf{D}_i der Standardabweichungen der Elemente von \mathbf{d}_i geschätzt werden:

$$(7-10) \quad \hat{\Sigma}^{(i)} = \mathbf{D}_i \mathbf{R} \mathbf{D}_i,$$

Anstatt die Forderung, dass in jeder Bedingung die Korrelationsmatrix gleich sein muss, zu überprüfen, wird der Empfehlung von Hedges und Olkin (1985, S. 209) nachgekommen und es werden für die Korrelationsmatrix \mathbf{R} die Interkorrelationen der Subtests der PBV genutzt. Da dann alle Versuchspersonen den Test in derselben Bedingung durchgeführt haben und die PBV immer noch als Referenz hinsichtlich der Äquivalenz dient, erscheint dieses Vorgehen angemessen. Da in vorliegender Arbeit $p - 1 = 10$ ist, liegt der kritische Wert für Q_i bei einem α von .10 bei 15.99. Kann die Homogenität der Effektstärken gezeigt werden, kann für eine Bedingung ein einziger Parameter bestimmt werden. „Die präziseste Schätzung der gemeinsamen Effektgröße δ_i^* ist gegeben durch

$$(7-11) \quad \hat{\delta}_i = d_{i1}w_{i1} + \dots + d_{ip}w_{ip} = \mathbf{w}_i' \mathbf{d}_i,$$

mit dem Gewichtungsvektor

$$(7-12) \quad \mathbf{w}_i = (w_{i1}, \dots, w_{ip})' = \frac{\Lambda^{(i)} \mathbf{e}}{\mathbf{e}' \Lambda^{(i)} \mathbf{e}} \text{ (Hedges \& Olkin, 1985, S. 212, übersetzt).}$$

$\hat{\delta}_i$ ist approximativ normalverteilt; die Varianz kann durch

$$(7-13) \quad \hat{\sigma}^2(\hat{\delta}_i) = \frac{1}{\mathbf{e}' \Lambda^{(i)} \mathbf{e}}$$

geschätzt werden, womit die Konstruktion eines Konfidenzintervalls möglich wird (vgl. Formel (7-3)). Beinhalten das 80 %-Konfidenzintervall nicht die Null, so kann mit einer Irrtumswahrscheinlichkeit von 10 % davon ausgegangen werden, dass der Effekt größer Null ist, womit von einem die Äquivalenz beeinflussenden Effekt ausgegangen werden müsste. Wird die Null durch das Intervall eingeschlossen, ist die Wahrscheinlichkeit einer gravierenden Beeinflussung eher gering; somit kann dann von der Gleichheit der Mittelwerte für diese Bedingung ausgegangen werden.

7.3 Vergleich der Varianzen

In einem zweiten Schritt wird für diese Fälle die Gleichheit der Varianz mittels T-2V-Test geprüft (nach Morgan, 1939, vereinfacht durch Diehl & Arbinger, 2001). Dieser Test ist zur Überprüfung der Gleichheit der Varianzen zwei abhängiger Stichproben konstruiert. Die Prüfgröße

$$(7-14) \quad |T_V| = \frac{|S_1^2 - S_2^2|}{\sqrt{\frac{4S_1^2 S_2^2 (1 - r_{12}^2)}{n-2}}}$$

ist t -verteilt mit $df = n - 2$ Freiheitsgraden. Das Signifikanzniveau wird mit $\alpha = .10$ etwas höher gewählt, da ja die Gleichheit der Varianzen bestätigt werden soll (Liebert & Liebert, 1995); Diehl und Arbinger (2001) empfehlen sogar $\alpha > .20$.

7.4 Analyse der Zusammenhänge

Neben der Gleichheit von Mittelwert und Varianz ist die Gleichheit der Reihenfolge der Probanden hinsichtlich der beiden Testungen, sprich die Korrelationen von PBV und CV, relevant. Da die Testwerte als intervallskaliert angenommen werden, wird der Produkt-Moment-Korrelationskoeffizient von Pearson (z. B. 1907) genutzt, der bekanntlich das Verhältnis von Kovarianz zu den Standardabweichungen der beiden betrachteten Variablen darstellt. Hedges und Olkin (1985) zeigen, wie analog zu den Effektstärken überprüft werden kann, ob die Korrelationen über die Subtests hinweg innerhalb einer Bedingung homogen sind und wie sie in diesem Fall entsprechend der Stichprobengröße gewichtet zusammengefasst werden können. Hier müssen die Korrelationskoeffizienten mittels Fisher's z-Transformation konvertiert werden: $z_i = .5[\ln(1 + r_i) - \ln(1 - r_i)]$, um r_i zu normalisieren. Der gewichtete Mittelwert der transformierten Korrelationen wird dann durch

$$(7-15) \quad z_+ = w_1 z_1 + \dots + w_p z_p \text{ mit } w_i = (n_i - 3) / \sum_{j=1}^p (n_j - 3)$$

berechnet. Analog zur Überprüfung der Homogenität der Effektstärken kann die $\chi^2(p - 1)$ -verteilte Statistik

$$(7-16) \quad Q = \sum_{i=1}^p (n_i - 3)(z_i - z_+)^2$$

zur Überprüfung der Homogenität der Korrelationen über die elf Subtests hinweg je Bedingung bestimmt werden, womit die Mittelung der Koeffizienten erst gerechtfertigt werden kann. Unter der Annahme der Normalverteilung von r kann dann entsprechend zu Formel (7-3) ein Konfidenzintervall für z_+ bestimmt werden:

$$(7-17) \quad (1 - \alpha) - \text{KI von } z_+ = z_+ \pm \frac{z(1-\alpha/2)}{\sqrt{\sum_{j=1}^p (n_j - 3)}}$$

Die Rückrechnung von z_i nach r_i erfolgt über die Formel $r_i = (e^{2z_i} - 1) / (e^{2z_i} + 1)$.

7.5 Gleichzeitiger Vergleich von Mittelwert, Varianz und Kovarianz

Für größere Stichproben ($n > 60$) besteht die Möglichkeit, mit Wilks (1946) L_{mvc} Mittelwerte, Varianzen und Kovarianzen von zwei bzw. prinzipiell k Testversionen gleichzeitig zu überprüfen:

$$(7-18) \quad L_{mvc} = \frac{|\Sigma|}{\bar{S}^2(1+(k-1)\bar{r}) \left[\bar{S}^2(1-\bar{r}) + \frac{1}{k-1} \sum_{i=1}^k (\bar{X}_i - \bar{X})^2 \right]^{k-1}}.$$

$|\Sigma|$ ist die Determinante der Kovarianzmatrix, \bar{S}^2 die mittlere Varianz und \bar{r} die mittlere Interkorrelation der Testversionen berechnet durch $\bar{r} = [k(k-1)\bar{S}^2]^{-1} \sum_{i \neq j=1}^k S_{ij}$. S_{ij} ist die Kovarianz zwischen Test i und Test j , die prinzipiell auch durch $S_{ij} = r_{ij} S_i S_j$ berechnet werden kann. Lienert und Raatz (1998) geben eine vereinfachte Formel von L_{mvc} für den Vergleich von zwei Tests an (S. 302):

$$(7-19) \quad L_{mvc} = \frac{S_1^2 S_2^2 (1 - r_{12}^2)}{\frac{S_1^2 + S_2^2}{2} (1 + r_{12}) \left[\frac{S_1^2 + S_2^2}{2} (1 - r_{12}) + \frac{1}{2} (\bar{X}_1 - \bar{X}_2)^2 \right]}.$$

Für größere Stichproben ist $-N \ln L_{mvc}$ approximativ χ^2 -verteilt mit $\frac{k}{2}(k+3) - 3$, also beim Vergleich von zwei Tests $df = 2$ Freiheitsgraden, wobei Wilks (1946) für $n > 60$ eine Genauigkeit von zwei Nachkommastellen angibt.

Der Vergleich von Varianzen und Kovarianzen ohne Berücksichtigung der Mittelwerte ist mittels Wilks L_{vc} möglich:

$$(7-20) \quad L_{vc} = \frac{|\Sigma|}{(\bar{S}^2)^k (1-r)^{k-1} (1+(k-1)r)}.$$

$-N \ln L_{vc}$ ist approximativ χ^2 -verteilt mit $\frac{k}{2}(k+1) - 2$, in diesem Fall mit $df = 1$ Freiheitsgraden. Für $k = 2$ Testversionen vereinfacht L_{vc} sich auf

$$(7-21) \quad L_{vc} = \frac{S_1^2 S_2^2 (1-r_{12}^2)}{\left(\frac{S_1^2 + S_2^2}{2}\right)^2 (1+r_{12})(1-r_{12})} = \frac{4S_1^2 S_2^2}{(S_1^2 + S_2^2)^2},$$

wobei hier besonders deutlich wird, dass $L_{vc} = 1$ ist, wenn die Varianzen gleich sind⁶.

Für den Fall, dass durch die zweite Testung eine Leistungssteigerung erfolgt, müssen für äquivalente Verfahren dennoch die Verteilungen gleich sein. Dies kann durch (7-21) überprüft werden.

⁶ Bei beispielsweise $S_2^2 = 2S_1^2$ resultiert $L_{vc} = \frac{8}{9} = .89$ und die $\chi^2(1)$ -verteilte Prüfgröße 0.12N; da der kritische Wert für $\alpha = .05$ bei einem Freiheitsgrad 3.841 ist, müsste somit $N > 33$ sein, damit der Unterschied als statistisch signifikant bezeichnet werden kann.

8. Testtheoretische Grundlagen

Die in den vorherigen Kapiteln formulierten Kriterien für Äquivalenz basieren auf den Annahmen der klassischen Testtheorie, wenngleich sie bei der Betrachtung der Fragestellung im Kontext anderer messtheoretischer Grundlagen (probabilistischer Testtheorien oder der Generalizability Theory von Cronbach, Gleser, Nande und Rajaratnam (1972)) ebenso Gültigkeit haben müssen. Lord (1980) legt in eindrucksvoller Kürze die Grundlagen der KTT dar. Diese Ausführungen liegen dem nächsten Abschnitt zugrunde.

8.1 True Score

Ausgehend von der Annahme, man hätte viele (n) parallele Versionen eines Tests, kann der wahre (Roh-)Wert (True Score, T) einer Person als Erwartungswert der Anzahl richtig gelöster Aufgaben durch den Mittelwert dieser n Testergebnisse (X_i) geschätzt werden. Aus dieser Definition ($T = E(X)$) können nun folgende Schlussfolgerungen abgeleitet werden:

Die Diskrepanz zwischen True Score und Messwert wird als Messfehler beschrieben:

$$(8-1) \quad E \equiv X - T.$$

Da T als Erwartungswert von X definiert ist, ist der Erwartungswert von E null:

$$(8-2) \quad \mu_{E|T} \equiv \mu_{(X-T)|T} \equiv \mu_{X|T} - \mu_{T|T} = T - T = 0.$$

Gleichung (8-2) zeigt, dass der Messfehler erwartungstreu ist. Ebenso ist der Mittelwert als Schätzer für den True Score erwartungstreu:

$$\mu_T \equiv \mu_{(X-E)} \equiv \mu_X - \mu_E = \mu_X.$$

Gleichung (8-2) kann als Regression von E auf T verstanden werden. Da $\mu_{(E|T)}$ unabhängig von T konstant 0 ist, hat auch diese Regressionsgleichung die Steigung null. Somit sind True Score und Messfehler unkorreliert:

$$(8-3) \quad \rho_{ET} = 0.$$

Die Varianz der Messwerte lässt sich nun aufgrund der Unkorreliertheit von E und T (daher gilt $\sigma_{TE} = 0$) wie folgt darstellen:

$$(8-4) \quad \sigma_X^2 \equiv \sigma_{T+E}^2 \equiv \sigma_T^2 + \sigma_E^2 + 2\sigma_{TE} = \sigma_T^2 + \sigma_E^2.$$

Die Kovarianz von X und T ist somit:

$$(8-5) \quad \sigma_{XT} \equiv \sigma_{(T+E)E} \equiv \sigma_T^2 + \sigma_{TE} = \sigma_T^2.$$

Die Testreliabilität, also die quadrierte Korrelation zwischen X und T (Gulliksen, 1950) lässt sich nach (8-5) somit bestimmen durch:

$$(8-6) \quad \rho_{XT}^2 \equiv \frac{\sigma_{XT}^2}{\sigma_X^2 \sigma_T^2} = \frac{\sigma_T^4}{\sigma_X^2 \sigma_T^2} = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_E^2}{\sigma_X^2}.$$

8.2 Paralleltests

Die Gleichungen (8-1) bis (8-6) können durch keine Stichprobe widerlegt werden. Sie erlauben aber auch keine Bestimmung von σ_T^2 , σ_E^2 oder ρ_{XT}^2 . Um diese Maße schätzen zu können, müssen nun Annahmen über die Zusammenhänge der Daten gemacht werden. Üblicherweise wird bei Anwendung der KTT angenommen, dass die Messfehler weder mit dem True Score noch untereinander korreliert sind. Für Test X und Y soll also gelten:

$$(8-7) \quad \rho_{E_X E_Y} = 0, \rho_{E_X T_Y} = 0, (X \neq Y)$$

Des Weiteren ist es üblich, die Reliabilität eines Testverfahrens durch zwei parallele Tests zu schätzen. Zwei Tests werden dann als parallel bezeichnet, wenn Probanden in beiden denselben True Score und dieselbe bedingte Fehlervarianz $\sigma_{E|T}^2$ haben (Lord & Novick, 1968):

$$(8-8) \quad T = T', \sigma_{E|T}^2 = \sigma_{E'|T'}^2$$

Aus den getroffenen zwei Annahmen lassen sich bei Vorliegen paralleler Tests True Score- und Messfehlervarianz schätzen und somit auch die Reliabilität:

$$(8-9) \quad \sigma_T^2 (= \sigma_{T'}^2) = \sigma_X^2 \rho_{XX'}$$

$$(8-10) \quad \sigma_E^2 (= \sigma_{E'}^2) = \sigma_X^2 (1 - \rho_{XX'})$$

$$(8-11) \quad \rho_{XT}^2 (= \rho_{X'T'}^2) = \rho_{XX'}$$

Eine Messung im Sinne der KTT stellt also die Schätzung des True Scores durch den Messwert dar, der im Falle von Leistungstests in der Regel die Anzahl richtig gelöster Aufgaben darstellt. Einen Nutzen der Betrachtung der Messfehlertheorie stellt Lord (1980) anhand eines Beispiels heraus: Ein Leistungstest wird einer größeren Anzahl von Kindern vorgelegt. Die Kinder mit den niedrigsten Punktzahlen werden für ein Training ausgewählt. Eine Woche später erfolgt die Überprüfung des Trainingserfolgs durch eine wiederholte Testung. Entsprechend der KTT gibt es zwei Gründe, warum ein Testwert niedrig sein kann: Entweder ist der True Score der Person niedrig oder der Betrag des Messfehlers hoch oder beides. Die Gruppe der Kinder mit den niedrigsten Werten haben am wahrscheinlichsten nicht nur niedrige True Scores, sondern auch hohe Messfehlerbeträge. Die Chance, dass bei einer zweiten Testung der Messfehler wieder hoch ausfällt, ist gering. Daher ist zu erwarten, dass diese Substichprobe im Mittel einen höheren Messwert bekommen wird, der ohne Berücksichtigung der KTT als Trainingseffekt erklärt werden würde. „This effect has caused many educational innovations to be mistakenly labeled ‚successful‘“ (Lord, 1980, S. 5).

Auch in den probabilistischen Testtheorien wird der Erwartungswert des Messwertes True Score genannt. Ebenso bezeichnet der Messfehler die Abweichung des Messwertes vom True Score. Daher sind die Messfehler notwendigerweise erwartungstreu und vom True Score unabhängig. Annahme (8-7) wird ebenso erfüllt wie alle anderen Gleichungen (Lord, 1980). Zum Schluss dieser Darstellung der KTT werden noch einige Schlussfolgerungen dargestellt, die spezifisch für die KTT sind.

8.3 Attenuitäts-Korrektur (correction for attenuation)

Die als praktische Anwendung durchaus umstrittene Minderungskorrektur der Korrelation zweier Merkmale hat in der KTT eine andere Implikation. Seien ξ und η der jeweilige True Score der Tests X und Y . Wie in (8-5) gilt $\sigma_{\xi\eta} = \sigma_{XY}$. Daher und wegen (8-10) gilt:

$$(8-12) \quad \rho_{\xi\eta} = \frac{\sigma_{\xi\eta}}{\sigma_{\xi}\sigma_{\eta}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y\sqrt{\rho_{XX'}\rho_{YY'}}} = \frac{\rho_{XY}}{\sqrt{\rho_{XX'}\rho_{YY'}}}.$$

Durch Umformen erhält man

$$(8-13) \quad \sqrt{\rho_{XX'}}\rho_{\xi\eta}\sqrt{\rho_{YY'}} = \rho_{XY}.$$

Da $0 \leq |\rho_{\xi\eta}\sqrt{\rho_{YY'}}| \leq 1$ muss wegen der zwingenden Verringerung des Produkts durch diesen Faktor

$$(8-14) \quad \sqrt{\rho_{XX'}} \geq \rho_{XY}$$

gelten, damit Gleichung (8-13) stimmen kann. Die Ungleichung (8-14) besagt, dass die Validität eines Tests bezüglich eines Kriteriums immer kleiner oder höchstens gleich der Wurzel der Reliabilität des Tests sein kann.

Gulliksen (1950) beschreibt die Bedeutung der Anwendung der correction for attenuation noch deutlicher. $\rho_{\xi\eta}$ wird nicht nur als „wahre Korrelation“ von Test und Kriterium beschrieben; Gulliksen benutzt dafür das Symbol $R_{\infty\infty}$, um deutlich zu machen, dass durch die Korrektur lediglich ein Wert angegeben wird, der zu erwarten wäre, würden Test und Kriterium mit unendlich vielen Items getestet. Zimmerman und Williams (1997) zeigen in einer Monte-Carlo-Studie, in welchen Fällen die Anwendung der correction for attenuation zu sinnvollen Ergebnissen führen kann: Nur bei Vorliegen hoher Reliabilität sowohl des Tests als auch des Kriteriums ($r > .8$) sowie einer relativ hohen Stichprobenzahl ($n > 80$) streuen die korrigierten Korrelationen in akzeptabler Weise um einen Mittelwert nahe dem True Score.

8.4 Testzerlegung

Wenn man sich einen Test aus einzelnen (parallelen) Subtests zusammengesetzt vorstellt, führt dies im Extremfall dazu, dass jedes Item als Subtest betrachtet werden kann. Der Testscore wird üblicherweise aus der Summe der „Subtests“ gebildet ($X = \sum_{i=1}^n Y_i$). Die Varianz von X lässt sich daher auch durch die Varianz der Subtests bestimmen:

$$(8-15) \quad \sigma_X^2 = \sum_{i=1}^n \sigma_{Y_i}^2 + \sum \sum_{i \neq j}^{n(n-1)} \sigma_{Y_i Y_j}.$$

Für die Reliabilität gilt dann entsprechend

$$(8-16) \quad \rho_{XX'} = \frac{\sum_i \sum_{i'} \sigma_{Y_i Y_{i'}}}{\sigma_X^2}.$$

Unter der Bedingung paralleler Subtests resultiert die Spearman-Brown-Formel:

$$(8-17) \quad \rho_{XX'} = \frac{n\rho_{YY'}}{1+(n-1)\rho_{YY'}}.$$

Aus (8-15) und (8-16) unter Berücksichtigung der Cauchy-Schwarz-Ungleichung (das Quadrat der Summe der Produkte von X_i und Y_i ist kleiner oder gleich der dem Produkt der Quadrate der

jeweiligen Summen X_i und Y_i) zeigt sich, dass der Koeffizient α (Cronbach, 1951) eine untere Grenze der Reliabilität darstellt:

$$(8-18) \quad \rho_{XT}^2 = \rho_{XX'} \geq \frac{n}{n-1} \left(1 - \frac{\sum \sigma_{Y_i}^2}{\sigma_X^2} \right) \equiv \alpha.$$

8.5 Reliabilitäts-Validitäts-Optimierungsparadoxon

Betrachtet man nun (8-18) und den daraus ableitbaren Validitätskoeffizienten

$$(8-19) \quad \rho_{XC} = \frac{\sum \sigma_{Y_i} \rho_{Y_i C}}{\sqrt{\sum_i \sum_j \sigma_{Y_i} \sigma_{Y_j} \rho_{Y_i Y_j}}},$$

wird deutlich, dass zur Maximierung der Reliabilität in (8-15) und (8-18) die Korrelation der Items untereinander möglichst hoch sein muss. Zur Maximierung der Validität (8-19) müssen die Items hingegen möglichst hoch mit dem Kriterium und möglichst gering untereinander korrelieren.

8.6 Itemanalyse

In der KTT bezeichnet das Verhältnis der richtigen Lösungen einer Aufgabe ($m_{r,i}$) in Bezug zu allen Antworten auf die gestellte Aufgaben (m) die Itemschwierigkeit ($p_i = m_{r,i}/m$). Je mehr Probanden eine Aufgabe richtig gelöst haben, desto leichter gilt sie. Dieser Wert ist jedoch offensichtlich abhängig von der getesteten Stichprobe. Gulliksen (1950) beschreibt hierzu einige Methoden, wie die Schwierigkeit einer Aufgabe in Abhängigkeit der getesteten Stichprobe transformiert werden kann. Ein weiterer Itemparameter ist die Trennschärfe, die als korrelativer Zusammenhang zwischen Itemwert und Testwert definiert ist (item-test-correlation). Obwohl „the biserial correlation *should not change* as the item difficulty changes with variations in group ability level“ (Gulliksen, 1950, S. 393), konnte dies bereits damals beobachtet (Carroll, 1945; Ferguson, 1941; Gulliksen, 1945; Richardson, 1936), aber nicht so recht erklärt werden. Aus der Formel für die punktbiseriale Korrelation, die zur Berechnung der Trennschärfe bei dichotomen Items dient:

$$(8-20) \quad r_{it(i)} = \frac{\bar{X}_{i+} - \bar{X}_{i-}}{s_x} \sqrt{p_i(1 - p_i)},$$

wobei \bar{X}_{i+} der Testmittelwert der Probanden, die Item i richtig beantwortet haben, \bar{X}_{i-} der Testmittelwert der Probanden, die Item i falsch beantwortet haben, s_x die Standardabweichung des Gesamttests und p_i die Itemschwierigkeit sind, wird jedoch deutlich, dass die Trennschärfe von der Itemschwierigkeit und damit ebenso von der Probandenstichprobe abhängt.

Exkurs: Lord (1980) zeigt, wie Trennschärfe a_i und Itemschwierigkeit b_i im Modell der probabilistischen Testtheorie (PTT) approximativ mit der Trennschärfe $r_{it(i)}$ der KTT zusammenhängen. Den Ausführungen liegen aber Bedingungen zugrunde, deren Aufführung zu viel Platz beanspruchen würde, weshalb nur die Quintessenz genannt werden soll: Die Anordnung der Items nach Schwierigkeit p_i ergibt eine andere Reihenfolge als die Anordnung nach Schwierigkeit b_i . Anders als in der KTT sind die Parameter der PTT wenigstens der Theorie nach invariant

gegenüber der Testung in verschiedenen Gruppen. Die Wahrscheinlichkeit, dass ein Proband eine bestimmte Aufgabe richtig löst, hängt nur von der Ausprägung seiner Fähigkeit θ ab, nicht aber von der Anzahl der Leute, die den Test mit gleicher Fähigkeitsausprägung bearbeitet haben und auch nicht von der Anzahl der Personen, die den Test bearbeitet haben und eine andere Fähigkeitsausprägung haben. Die Schwierigkeit eines Items b_i bleibt gleich, unabhängig davon, wie fähig die Probanden der Stichprobe sind. In der KTT ist die Schwierigkeit eines Items in einer fähigen Gruppe geringer als in einer weniger fähigen. Ebenso ist eine Fünfkilohantel für einen Bodybuilder leichter als für einen Doktoranden der Psychologie, obwohl das Gewicht ja immer gleich ist.

9. Untersuchung 1: Vortestung an Studenten

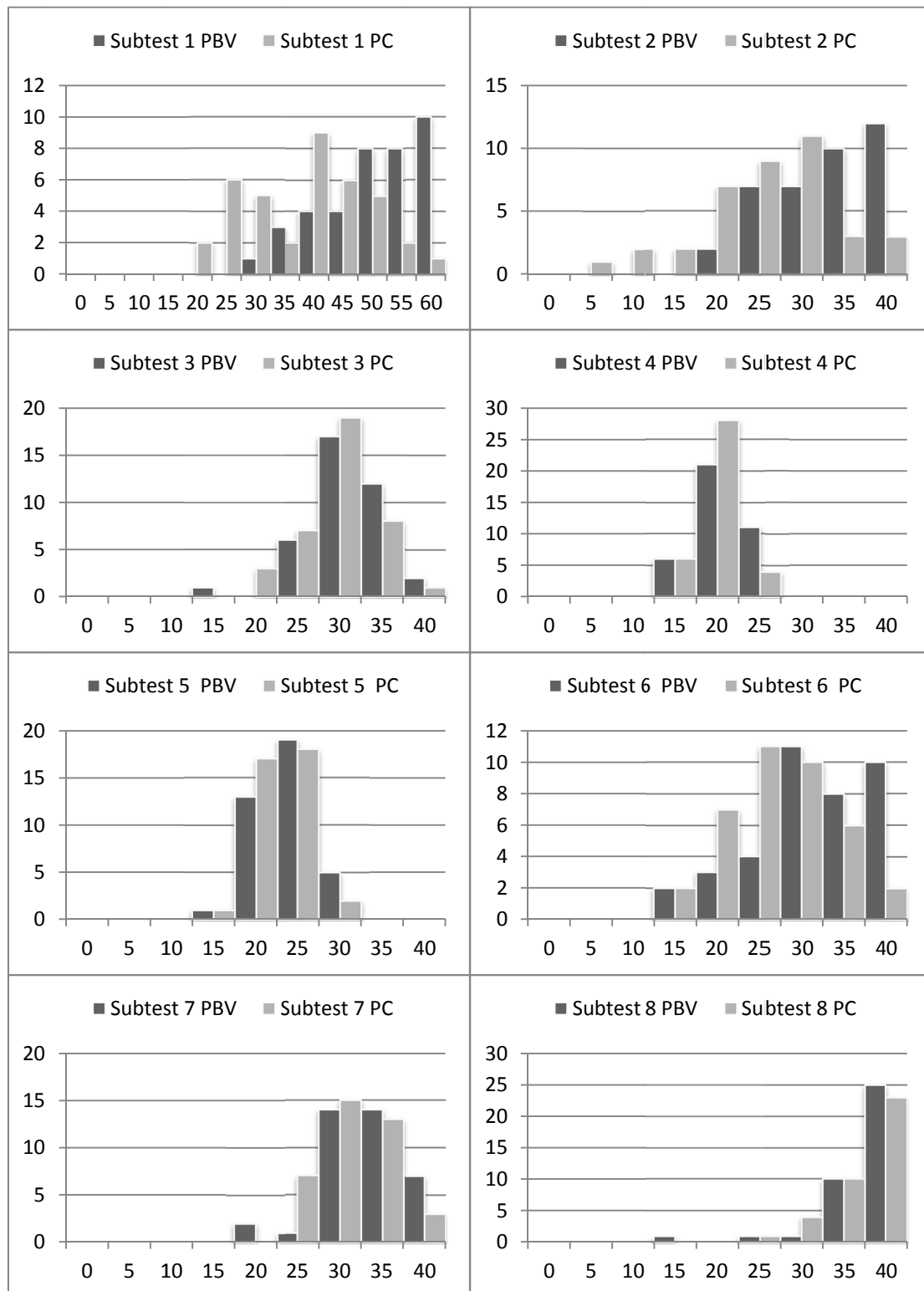
Zur Überprüfung der Funktionalität der Computerversion (CV), des neu designten Fragebogens und der psychometrischen Eigenschaften der Subtests sowie der gewählten Zeitvorgaben wurde die überarbeitete Version des LPS an einer studentischen Stichprobe getestet. In Tabelle 9.1 werden Stichprobengröße und die Altersstruktur der $n = 38$ Studierenden dargestellt.

Tabelle 9.1: Stichprobengröße und die Altersstruktur der studentischen Stichprobe (Untersuchung 1)

	PBC - CV		CV - PBV	
	weiblich	männlich	weiblich	männlich
n	10	5	19	4
	15		23	
	38			
<i>M</i> (<i>SD</i>)	24;6 (2;7)	22;11 (1;1)	24;8 (3;11)	24;11 (0;11)
	24;0 (2;3)		24;8 (3;7)	
	24;5 (3;1)			

Bemerkung: Altersangaben in der Form Jahre;Monate

Die Teilstichprobe, die zuerst die CV durchführte, wurde zwischen März und April 2007 in einer Einzeltestung in einem Labor an der Universität unter standardisierten Bedingungen getestet. Im Mai 2007 wurden in vier Gruppentestungen alle Probanden mittels der Papier-Bleistift-Version (PBV) untersucht. Die zweite Gruppe führte die CV Ende Juni und Anfang Juli in einer unkontrollierten Bedingung aus, indem sie mittels eines Zugangscodes den Tests in den Wochen nach der Testung mittels der PBV an einem beliebigen PC ausfüllen konnten. Die Zeit zwischen den beiden Testungen schwankte dabei insgesamt sehr stark zwischen drei und 74 Tagen und liegt im Mittel bei 40 ($SD = 18$) Tagen.



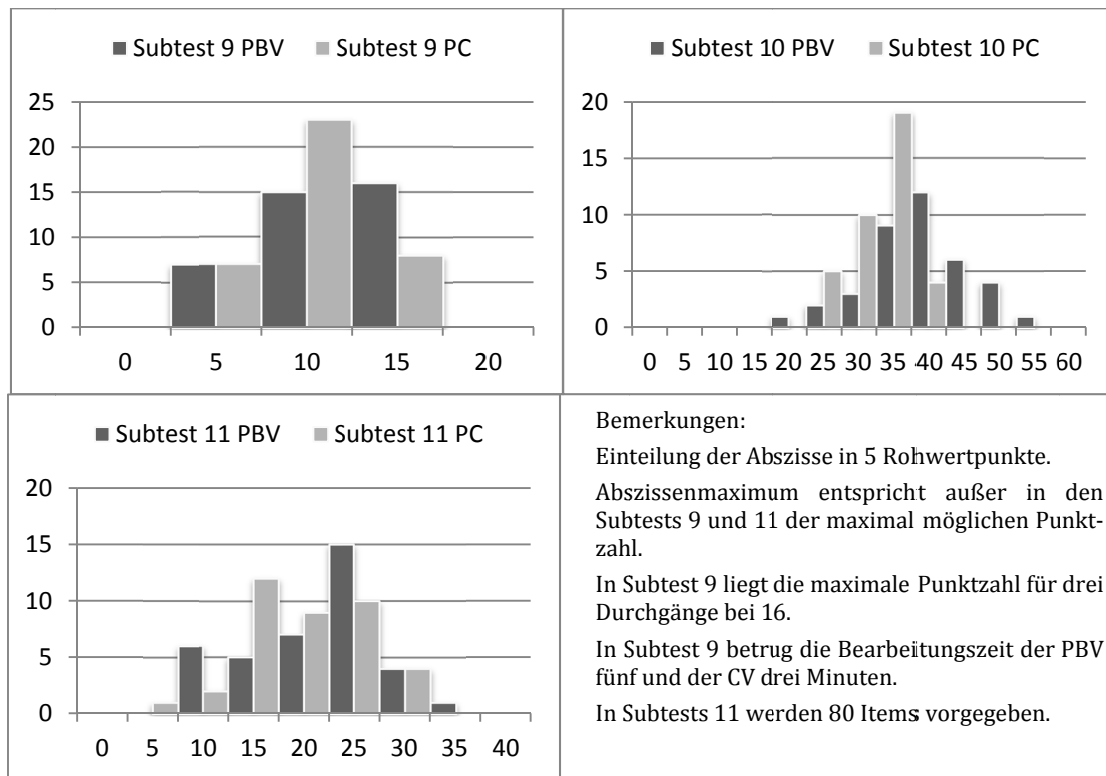


Abbildung 9.1: Histogramme der elf Subtests der Papier-Bleistift-Version und der Computer-version der studentischen Stichprobe unter Berücksichtigung der vorgeschlagenen Zeitvorgaben

9.1 Papier-Bleistift-Testung der studentischen Stichprobe

Die 38 studentischen Probanden wurden in vier Gruppen mittels der neu erstellten PBV des LPS-neu getestet. Hierbei wurden ihnen für die Bearbeitung der Aufgaben je Subtest zunächst die in Tabelle 2.32 (Seite 30) vorgeschlagenen Zeiten zur Verfügung gestellt. Nach Ablauf dieser Zeit wurden die Studenten gebeten, einen Strich hinter der zuletzt beantworteten Aufgabe zu ziehen und anschließend die restlichen Aufgaben des Subtests nach der Markierung zu bearbeiten. Es zeigte sich, dass die Aufgaben von Subtest 4 und 5 insgesamt so schwierig waren, dass kein Student sie innerhalb der maximal vorgegebenen Zeit von zehn Minuten alle lösen konnte. Die hohe Anzahl der Aufgaben in Subtest 11 gründet auf dem Umstand, dass damit zusätzlich länger andauernde Konzentrationsleistung erhoben werden soll. Damit stellt sich für diesen und ebenso für Subtest 9 die Frage nach einer Gesamtbearbeitung nicht. Die minimalen und die maximalen Bearbeitungszeiten werden in Tabelle 9.2 aufgelistet.

Die Verteilungen für die Subtests 1, 2, 6 und 8 der PBV (vgl. Abbildung 9.1) zeigen, dass die Aufgaben für eine studentische Stichprobe tendenziell zu leicht sind. Ein zu großer Anteil konnte innerhalb der Zeit alle Items teilweise sogar fehlerfrei bearbeiten (siehe auch Tabelle 9.2). In den übrigen Subtests lässt sich – eingeschränkt durch die geringe Fallzahl und teilweise durch eine geringe Varianz – prinzipiell eine Normalverteilung ausmachen. Genauere statistische Analysen bezüglich Schiefe und weiterer Verteilungsmaße, ebenso in Hinblick auf die Itemkennwer-

te, werden aufgrund der Stichprobengröße ausgespart und nur für die Hauptuntersuchung berechnet.

Tabelle 9.2: Bearbeitungsmerkmale der PBV der vier Gruppen der studentischen Stichprobe

Subtest	t	Min	Max	item _(min)	<i>n_{alles}</i>
1	180	90	350	26	19
2	120	140	450	20	8
3	240	210	490	17	** Nr. 39
4	240	-	*600	0	** Nr. 27
5	240	-	*600	0	** Nr. 32
6	120	100	330	19	7
7	180	170	350	26	9
8	180	80	310	23	24
9	300	-	*300	-	-
10	120	150	310	20	1
11	300	-	*300	-	-

Bemerkungen: Zeitangaben in Sekunden; t: geplante Zeitvorgabe; item_(min): geringste Anzahl bearbeiteter Items; *n_{alles}*: Anzahl der Vpn, die alle Items bearbeiteten; * maximale Bearbeitungszeit des Subtests; ** keine Vp schaffte alle Items, die Angabe steht für das letzte Item, das von einer Vp bearbeitet wurde

9.2 Computer-Testung der studentischen Stichprobe

Für den Vorversuch wurden nicht alle in Kapitel 5.1 formulierten Bedingungen variiert. Die Items wurden alle nur einzeln dargestellt, da sich damit die Zahl der Bedingungen prinzipiell halbiert. Es resultierte somit ein $1 \times 3 \times 3$ -Design, das in Tabelle 9.3 veranschaulicht wird. Bei der Durchführung wurde die Reihenfolge der Bedingungen konstant gehalten. Dafür änderte sich je nach Versuchspersonennummer der erste zu bearbeitende Subtest. Die Versuchspersonen 1, 12, 23 und 34 begannen mit Subtest 1, Versuchspersonen 2, 13, 24 und 35 mit Subtest 2 usw.; mit den Subtests 9 und 10 wurde wegen ihres Designs, das nicht für die Zeitmessung bei einzelnen Items geeignet ist, nicht begonnen, weshalb die Versuchspersonen 9, 10, 11, 20, 21, 22 usw. mit Subtest 11 begannen. Für die Subtests 9 und 10 wurde die Bedingung stattdessen über alle Versuchspersonen konstant gehalten. Da beide Verfahren Items beinhalten, bei denen keine Antwort möglich sein muss, wurde das Antwortformat „weiter“ gewählt. Es gab also einen „weiter“-Button, der nach der Auswahl einer Lösung geklickt werden musste. Auch die Beschränkung der Bearbeitungszeit einzelner Items macht für diese Subtests keinen Sinn. Daher wurde die Bearbeitungszeit für den Gesamttest vorgeben.

Tabelle 9.3: Übersicht der Bedingungen der Präsentation der Subtests der Computerversion in der studentischen Stichprobe

Präsentation		Einzel								
Antwortformat		Einmalig			Weiter			Korrigierbar		
Zeitvorgabe		Ohne	Gesamt	Item	Ohne	Gesamt	Item	Ohne	Gesamt	Item
Bedingung		1	2	3	4	5	6	7	8	9
Gruppe	1	1	2	3	4	5	6	7	8	11
	2	2	3	4	5	6	7	8	11	1
	3	3	4	5	6	7	8	11	1	2
	4	4	5	6	7	8	11	1	2	3
	5	5	6	7	8	11	1	2	3	4
	6	6	7	8	11	1	2	3	4	5
	7	7	8	11	1	2	3	4	5	6
	8	8	11	1	2	3	4	5	6	7
	9	11	1	2	3	4	5	6	7	8

Bemerkungen: Die Subtests 9 und 10 wurden für alle Gruppen im Antwortformat „weiter“ und mit Zeitvorgabe „Gesamt“ präsentiert.

Jede Versuchsperson hatte für die Bearbeitung der Subtests 1 bis 8 jeweils maximal 10 Minuten Zeit und wurde in der Instruktion darüber informiert. Vor jedem Subtest erhielten die Versuchspersonen eine der Bedingung angepasste Instruktion, die auf den ausführlichen Angaben der Instruktion der PBV basierte (vgl. Anhang 15.2). Die Möglichkeiten, Aufgaben auszulassen, zu wiederholen und dgl., waren aufgrund der unterschiedlichen Bedingungen der CV am PC teilweise größer, teilweise eingeschränkter, so dass ein direkter Vergleich mit denselben Angaben wie in Tabelle 9.2 nicht sinnvoll ist. Auch hier wurden jedoch in den Subtests 4, 5, 9 und 11 von keinem Probanden alle Items bearbeitet. Die Umsetzung von Subtest 9 gelang nicht wie gewünscht. Die kurzen Bearbeitungszeiten (vgl. Tabelle 9.4) sind einmal bedingt durch einen Programmierfehler, der die Zeit anstatt auf fünf auf drei Minuten begrenzte; zum anderen wurde von einigen Probanden der Subtest nach einem Durchgang (jede achte „Null“ zählen) offensichtlich irrtümlich beendet.

Subtest 10 war auf vier Minuten begrenzt, da sich bei der Konstruktion gezeigt hatte, dass eine längere Zeitvorgabe eher langsames Arbeiten zur Folge hatte. Während der Erstellung wurden auch nie annähernd vier Minuten benötigt. Dennoch erwies sich dieses Limit für 22 Probanden als zu niedrig. Andererseits beendeten zehn Versuchspersonen den Test vorzeitig vollständig bearbeitet, die restlichen sechs Probanden hatten Subtest 10 ebenfalls vollständig bearbeitet, allerdings die Bearbeitungszeit vollständig genutzt bzw. abgewartet. Da für Subtest 11 ohnehin davon ausgegangen wird, dass die 80 Items auch in 10 Minuten nicht alle gelöst werden können, wurde dieser Subtest von vornherein auf fünf Minuten begrenzt.

Tabelle 9.4: Bearbeitungszeiten der CV der vier Gruppen der studentischen Stichprobe

Subtest	t	Min	Max	M	SD
1	180	139	484	298	90
2	120	90	*607	302	99
3	240	134	581	307	95
4	240	73	*607	552	116
5	240	308	*607	530	89
6	120	115	310	194	46
7	180	178	510	297	83
8	180	97	440	208	83
9	**180	73	186	145	46
10	120	165	242	234	16
11	300	293	323	301	4

Bemerkungen: Zeitangaben in Sekunden; t: geplante Zeitvorgabe; Wertungenauigkeiten aufgrund von Rundungen; * systembedingtes Maximum von 607 Sekunden (10 Minuten); ** wegen eines Tippfehlers in der Programmierung ist die Bearbeitungszeit in der CV nur drei anstatt fünf Minuten

Die Verteilungen der Subtests der CV weisen eine deutlich größere Ausgewogenheit auf als in der PBV; eine Ausnahme stellt Subtest 8 dar, der in der CV ebenso eindeutig zu leicht für die Versuchspersonen gewesen zu sein scheint (vgl. Abbildung 9.1). Aufgrund der unbeabsichtigten Zeitverkürzung bei Subtest 9 überrascht die relative Symmetrie. Die Ergebnisse von Subtest 10 variieren nicht so stark; auch hier kann die Zeitbegrenzung als mögliche Ursache angeführt werden.

9.3 Vergleich der Durchführungsmodi und Konsequenzen der Vorstudie

Bezüglich der Bearbeitungszeiten zeigte sich, dass die angepassten Vorgaben in den meisten Fällen angemessen erscheinen. Unter Berücksichtigung des Umstandes, dass die Probanden Studenten waren, deren Fähigkeiten in allen getesteten Bereichen zumindest vorwiegend überdurchschnittlich sein sollten, ist der teilweise hohe Anteil derer, die die Subtests innerhalb der vorgegeben Zeit absolvierten, erwartbar. Die Subtests 4 und 5 stellen sich allerdings insgesamt als zu schwierig heraus, um in der vorgegeben Zeit vollständig bearbeitet werden zu können, weshalb die Bearbeitungszeit auf fünf Minuten verlängert wird. Zum Ausgleich wurde die Bearbeitungsdauer von Subtest 3 auf drei Minuten reduziert.

Die direkte Vergleichbarkeit der beiden Modalitäten (Tabelle 9.5) war nicht Ziel der Vorstudie. Eine Gegenüberstellung ist insofern schwer interpretierbar, als dass die unterschiedlichen Bedingungen der CV verschiedene Effekte zur Folge haben können. Insbesondere die Korrelationen der beiden Verfahren zeigen jedoch, dass zumindest bedingt ähnliche Reihenfolgen der Versuchspersonen resultieren. Die statistisch signifikanten Mittelwertsunterschiede der Subtests 1, 2, 6 und 10 weisen darauf hin (ebenso wie alle anderen Mittelwertsdifferenzen), dass sich die PBV für die Mehrheit der Studenten als leichter erwiesen hatte. Der Effekt ist lediglich für Subtest 8 gering. In allen anderen Subtests unterscheiden sich PBV und CV mittel bis stark.

Tabelle 9.5: Mittelwerte und Standardabweichungen der elf Subtests der Papier-Bleistift-Version und der Computer-Version der studentischen Stichprobe, bivariate Korrelation, Mittelwertsdifferenzen, Effektstärke und z-Tests der korrespondierenden Subtests

	PBV		CV		Vergleich						
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>r</i>	<i>p</i>	<i>M(D)</i>	<i>SD(D)</i>	<i>d</i>	<i>z</i>	<i>p</i>
Subtest 1	48.42	8.20	36.63	10.62	.32	.05	11,79	11.14	1.06	5.42	<.01
Subtest 2	31.05	5.90	23.74	7.81	.32	.05	7,32	8.14	0.90	4.61	<.01
Subtest 3	29.16	4.69	27.89	4.76	.54	<.01	1,26	4.51	0.28	1.17	.25
Subtest 4	18.76	2.75	17.89	2.91	.41	.01	0,87	3.09	0.28	1.34	.19
Subtest 5	21.50	3.48	20.39	2.89	.40	.01	1,11	3.52	0.31	1.51	.14
Subtest 6	29.39	7.35	24.89	6.13	.45	.01	4,50	7.14	0.63	2.90	<.01
Subtest 7	31.26	4.54	29.53	3.87	.45	.01	1,74	4.47	0.39	1.79	.08
Subtest 8	36.16	5.31	35.39	4.38	.54	<.01	0,76	4.73	0.16	0.68	.50
Subtest 9	9.00	3.52	7.87	2.64	.27	.11	1,13	3.79	0.30	1.59	.12
Subtest 10	37.42	7.61	30.89	4.15	.33	.04	6,53	7.35	0.89	4.64	<.01
Subtest 11	19.53	6.56	18.00	6.07	.66	<.01	1,53	5.23	0.29	1.05	.30

In Abbildung 9.2 wird deutlich, dass die verschiedenen Bedingungen der CV zu einer gewissen Bandbreite an unterschiedlichen Ergebnissen führt, die bei einigen Subtests unter der durchschnittlichen Leistung in der PBV liegen (Subtests 1, 2, 6 und 10), bei einigen aber auch um die durchschnittliche Leistung der PBV streuen (3, 4, 5, 7, 8, 9, 11).

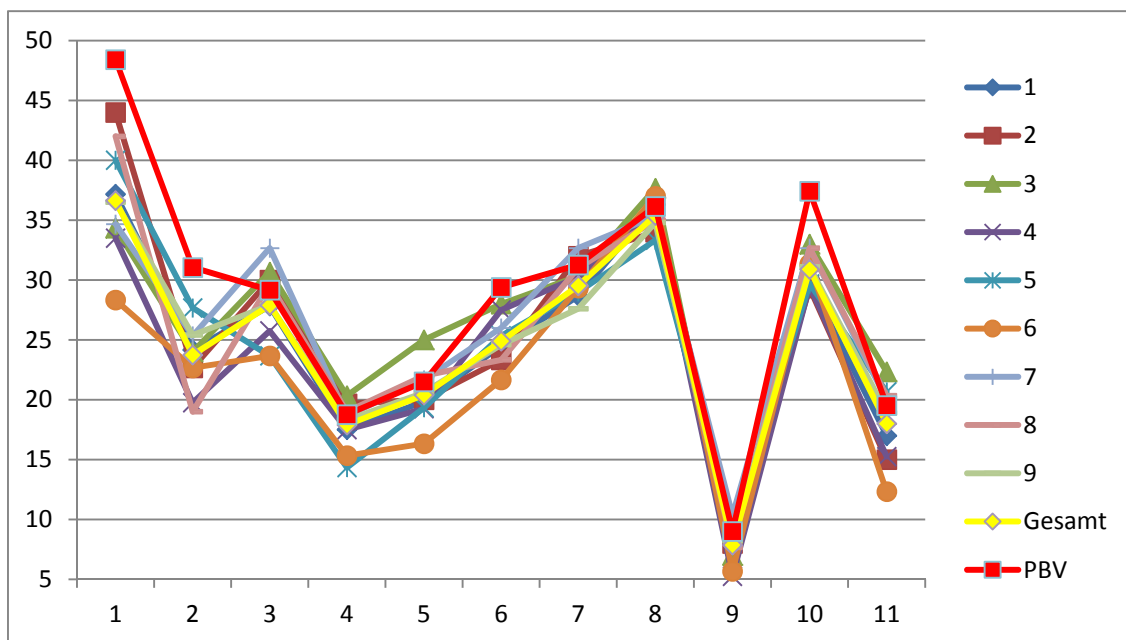


Abbildung 9.2: Mittelwerte der Subtests und des Gesamtwerts der CV in den unterschiedlichen Bedingungen (vgl. Tabelle 9.3) sowie die Mittelwerte der entsprechenden PBV

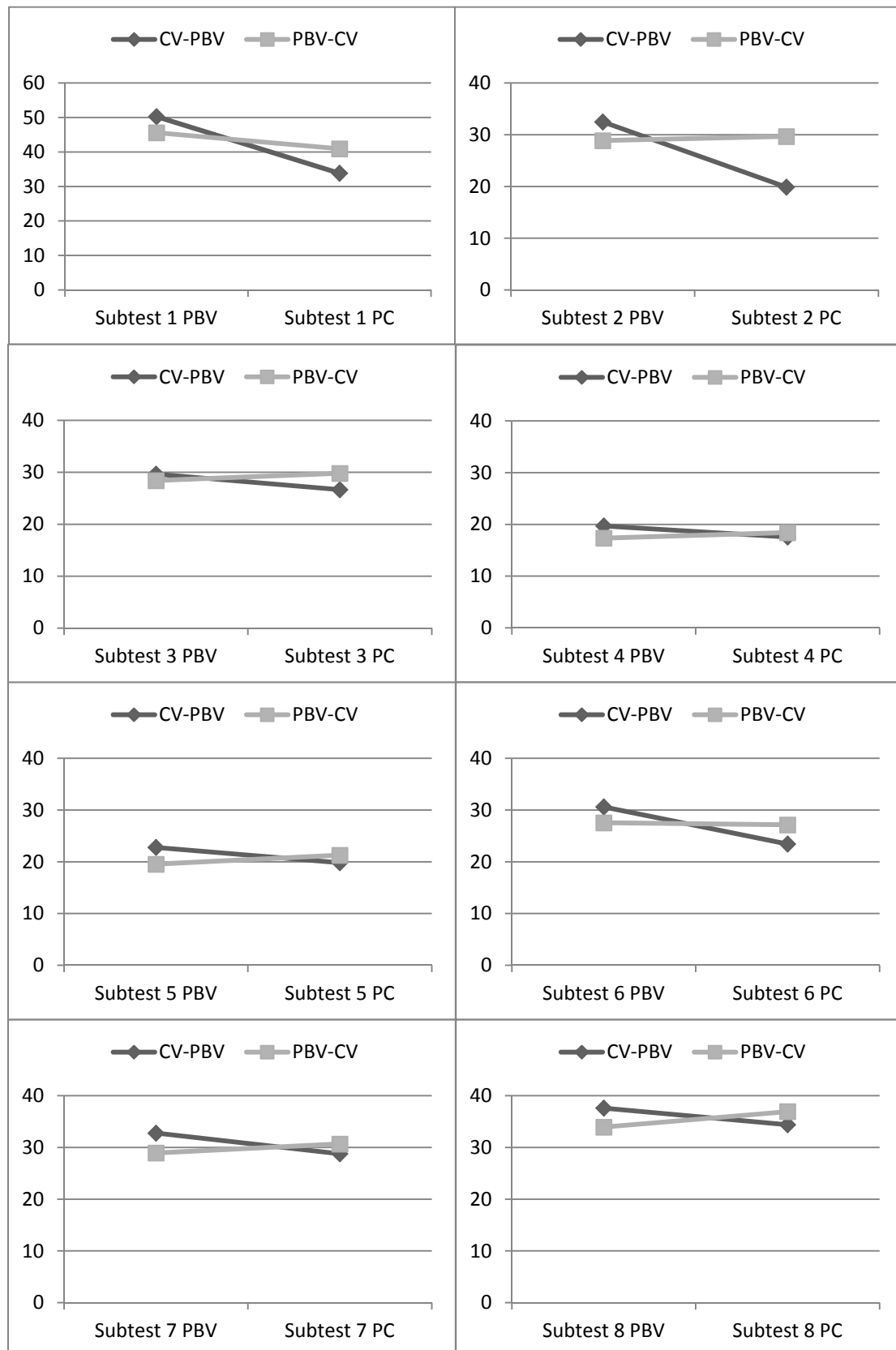
Die in Tabelle 9.6 dargestellten Zusammenhänge der Subtestergebnisse mit der Abiturnote zeigen, dass das LPS-neu bivariat nur geringe Erklärungskraft bezüglich schulischer Leistungen aufweist. Die Subtests 1 und 2, die Allgemeinbildung abbilden sollen, korrelieren sogar entgegen

der erwarteten Richtung. Der Gesamtscore korreliert nur gering negativ. Als beste Prädiktoren erweisen sich logisches Denken (Subtests 3, 4 und 5) und räumliche Wahrnehmung (Subtest 7 und 8). Es ist dabei jedoch zu berücksichtigen, dass die Abiturnoten von Psychologiestudierenden relativ homogen sind ($M = 1.73$, $SD = 0.49$). Für die CV gilt zudem, wie beim Vergleich mit der PBV, dass die unterschiedlichen Bedingungen verschiedene Effekte und somit Verzerrungen aufweisen können. Andererseits sind die Korrelationen von PBV und CV mit der Abiturnote ähnlich, was zumindest als Indikator für externe und Konstruktvalidität gewertet werden kann.

Tabelle 9.6: Bivariate Korrelation der elf Subtestergebnisse und des Gesamtscores der Papier-Bleistift-Version, der Computer-Version und des Mittelwertes beider Testungen der studentischen Stichprobe mit der Abiturnote ($n = 38$)

	PBV		CV		$M_{PBV\&CV}$	
	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
Subtest 1	.16	.34	.28	.09	.28	.09
Subtest 2	.17	.32	.06	.72	.13	.44
Subtest 3	-.31	.06	-.42	.01	-.41	.01
Subtest 4	-.16	.35	-.10	.55	-.15	.36
Subtest 5	-.38	.02	-.37	.02	-.45	<.01
Subtest 6	.00	.99	-.08	.65	-.04	.81
Subtest 7	-.25	.14	-.34	.04	-.34	.04
Subtest 8	-.52	<.01	-.47	<.01	-.56	<.01
Subtest 9	.02	.89	-.20	.23	-.09	.59
Subtest 10	.01	.98	-.31	.06	-.13	.45
Subtest 11	-.05	.78	-.14	.41	-.10	.55
Gesamt	-.13	.43	-.18	.28	-.18	.27

In der Vorstudie wurde die Reihenfolge der bearbeiteten Modi variiert. 23 Studenten bearbeiteten zuerst die CV, 15 Studenten zuerst die PBV (vgl. Tabelle 9.1). Zur Überprüfung der Effekte bezüglich der Reihenfolge und des Modus wurde eine 2×2 Varianzanalyse mit Messwiederholung berechnet. In Tabelle 9.7 wird die deskriptive Statistik dargestellt. Bereits hier werden einige Effekte augenscheinlich. In Abbildung 9.3 werden die Unterschiede grafisch veranschaulicht. Das linke helle Quadrat entspricht der PBV von Versuchspersonen, die zuerst die PBV bearbeitet haben, das rechte Quadrat deren CV. Entsprechend wird die erste Testung der Versuchspersonen, die zuerst die CV bearbeitet haben, durch die rechte dunkle Raute dargestellt. Chronologisch sind die dunklen Werte also von rechts nach links zu lesen.



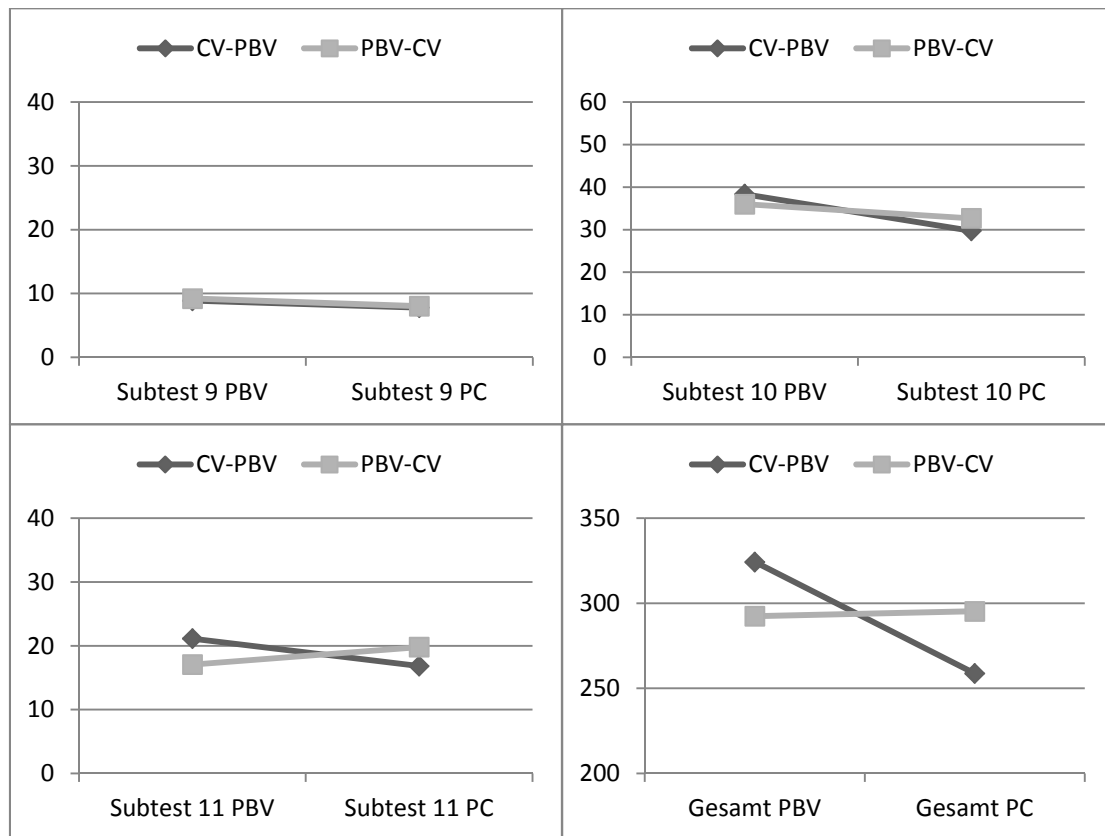


Abbildung 9.3: Diagramme der Mittelwerte der elf Subtests und des Gesamtwerts des LPS-neu der studentischen Stichprobe aufgeschlüsselt nach Modus und Reihenfolge

In den Diagrammen werden bei jedem Subtest (außer 9) und dem Gesamtwert Interaktionseffekte sichtbar (vgl. Tabelle 9.8). Für die Subtests 2, 3, 4, 5, 6, 7 und 11 ist diese Interaktion so zu interpretieren, dass die Versuchspersonen in der zweiten Testung eine bessere Leistung zeigen als in der ersten, unabhängig davon, in welchem Modus die erste Bearbeitung stattfand. Bei den Subtests 1, 6 und 10 liegen hingegen deskriptiv Verschlechterungen der Leistung der CV vor, wenn die Probanden zuerst die PBV bearbeiteten. Die Interaktion begründet sich hierbei darin, dass die Gruppe, die mit der CV begann, sich bei der zweiten Testung deutlich verbesserte, sich in der anderen Gruppe hingegen nur geringe Änderungen ergaben. Insgesamt können also bessere Leistungen in der PBV identifiziert werden. Die leistungssteigernden Effekte einer zweiten Testung werden bei vier Subtests durch leistungsreduzierende Effekte der CV kompensiert.

Tabelle 9.7: Mittelwerte und Standardabweichungen der elf Subtests und des Gesamtwerts des LPS-neu der studentischen Stichprobe getrennt nach Modus und Reihenfolge

Reihenfolge	PBV-CV ($n = 15$)				CV-PBV ($n = 23$)			
Modus	PBV		CV		PBV		CV	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Subtest 1	45.60	7.80	40.93	9.68	50.26	8.09	33.83	10.45
Subtest 2	28.87	6.32	29.67	5.56	32.48	5.27	19.87	6.57
Subtest 3	28.40	4.81	29.80	4.86	29.65	4.65	26.65	4.35
Subtest 4	17.33	2.13	18.40	2.77	19.70	2.74	17.57	3.01
Subtest 5	19.53	2.45	21.27	3.41	22.78	3.49	19.83	2.41
Subtest 6	27.53	7.13	27.13	6.57	30.61	7.38	23.43	5.49
Subtest 7	28.93	5.12	30.67	4.59	32.78	3.45	28.78	3.22
Subtest 8	33.93	6.22	36.93	4.57	37.61	4.15	34.39	4.03
Subtest 9	9.20	3.51	8.00	2.42	8.87	3.60	7.78	2.83
Subtest 10	36.00	6.97	32.67	3.52	38.35	8.01	29.74	4.19
Subtest 11	17.07	5.50	19.80	6.56	21.13	6.80	16.83	5.56
Gesamt	292.40	35.21	295.27	31.36	324.22	33.84	258.70	30.03

Es liegt kein Haupteffekt für den Faktor Reihenfolge vor. Über beide Testungen hinweg unterscheiden sich die beiden Gruppen also nicht voneinander. Für Subtest 9 liegt ebenfalls keine Interaktion der beiden Faktoren vor. Für diesen Subtest zeigte sich lediglich, dass in der PBV etwas mehr Items gelöst wurden als in der CV.

Die signifikanten Haupteffekte des Faktors Modus (Subtests 1, 2, 6, 7, 9, 10 und Gesamt) begründen sich darin, dass die Versuchspersonen in der PBV im Mittel eine bessere Leistung zeigten als in der CV. Bei Berücksichtigung der Interaktionen zeigt sich somit sowohl ein leistungsmindernder Einfluss der CV als auch eine Leistungssteigerung bei der zweiten Testung als PBV. In den Subtests 3, 4, 5, 8 und 11 scheint der leistungsförderliche Effekt der zweiten Testung hingegen in beiden Modi zu wirken; der leistungsmindernde Effekt der CV tritt hier nicht bzw. in sehr begrenztem Maße auf.

Tabelle 9.8: *F*-Wert, Signifikanzniveau und partielles und klassisches η^2 der 2×2 Varianzanalyse mit Messwiederholung bezüglich der Faktoren Modus (within) und Reihenfolge (between) der elf Subtests und des Gesamtwerts des LPS-neu der studentischen Stichprobe ($n = 38$)

	Modus				Reihenfolge				Modus × Reihenfolge			
	<i>F</i>	<i>p</i>	$\eta^2(M)$	η^2	<i>F</i>	<i>p</i>	$\eta^2(R)$	η^2	<i>F</i>	<i>p</i>	$\eta^2(M \times R)$	η^2
Subtest 1	43.66	<.01	.55	.23	0.23	.64	.01	<.01	13.58	<.01	.27	.07
Subtest 2	55.45	<.01	.61	.15	2.92	.10	.07	.04	71.50	<.01	.67	.20
Subtest 3	1.45	.24	.04	.01	0.47	.50	.01	.01	10.96	<.01	.23	.05
Subtest 4	1.43	.24	.04	.01	0.94	.34	.03	.02	12.87	<.01	.26	.08
Subtest 5	1.88	.18	.05	.01	1.04	.31	.03	.02	27.66	<.01	.43	.13
Subtest 6	12.76	<.01	.26	.07	0.03	.87	<.01	<.01	10.21	<.01	.22	.06
Subtest 7	3.81	.06	.10	.02	0.68	.42	.02	.01	24.37	<.01	.40	.11
Subtest 8	0.03	.86	<.01	<.01	0.16	.69	<.01	<.01	26.44	<.01	.42	.10
Subtest 9	3.21	.08	.08	.03	0.11	.74	<.01	<.01	0.01	.93	<.01	<.01
Subtest 10	26.68	<.01	.43	.19	0.03	.86	<.01	<.01	5.21	.03	.13	.04
Subtest 11	1.44	.24	.04	<.01	0.08	.78	<.01	<.01	28.81	<.01	.44	.08
Gesamt	114.73	<.01	.76	.15	0.05	.82	<.01	<.01	136.69	<.01	.79	.18

Es zeigt sich also, dass äquivalente Bedingungen nicht a priori angenommen werden können und dass der Bearbeitungsmodus in aller Regel eine Rolle spielt. Andererseits besteht, wie aus Abbildung 9.2 ersichtlich, die Chance, durch die Variation der Bedingungen der CV zwei äquivalente Durchführungsmodalitäten zu finden. Aufgrund der Anmerkungen der Studierenden wurden kleine Änderungen bei einzelnen Items des LPS-neu vorgenommen. Insgesamt schien aber die Neufassung hinsichtlich der Durchführbarkeit gelungen. Die unterschiedliche Bearbeitungsreihenfolge der Subtests irritierte ein wenig bei der zweiten Testung, weshalb in der Hauptuntersuchung das LPS-neu in beiden Modi mit Subtest 1 begonnen wurde. Auch wenn die Bearbeitungszeiten einiger Subtests etwas zu lang erschienen, wurde beschlossen, die meist originalen Zeitvorgaben beizubehalten und eine diesbezügliche Revidierung basierend auf den Ergebnissen der Hauptuntersuchung, sofern nötig, vorzunehmen.

10. Untersuchung 2: Testung an Berufsschülern

10.1 Design

Die Hauptuntersuchung wurde an der Städtischen Berufsschule I Regensburg - Berufliches Schulzentrum Georg Kerschensteiner durchgeführt⁷. Insgesamt nahmen 298 Schülerinnen und Schüler an den Testungen teil. Die Papier-Bleistift-Version (PBV) bearbeiteten insgesamt 279, vollständig 273 Schüler; sechs Schüler gaben unvollständige, nicht verwertbare oder offensichtlich nicht im Sinne der Testung bearbeitete Testbögen ab. Von 273 Schülern wurde die Computerversion (CV) gestartet, 223 bearbeiten sie vollständig. Einige Schüler beendeten die Testung vorzeitig (was jedem Teilnehmenden offen stand), einige Schüler ließen einzelne Subtests aus, indem sie einen Subtest beendeten, ohne bei einem einzigen Item eine Lösung angeklickt zu haben. 205 Schüler bearbeiteten beide Versionen vollständig (vgl. Tabelle 10.1).

Tabelle 10.1: Anzahl der bearbeiteten Subtests

	0*	4	5	6	7	8	9	10	11	Vpn	Fehlend	Gesamt
PBV	6 (2.01)	-	-	-	-	-	-	-	273 (91.61)	279 (93.62)	19 (6.38)	298 (100)
CV	2 (0.67)	1 (0.34)	1 (0.34)	1 (0.34)	2 (0.67)	11 (3.69)	5 (1.68)	27 (9.06)	223 (74.83)	273 (91.61)	25 (8.39)	298 (100)
PBV & CV										205 (68.79)	93 (31.21)	298 (100)

Bemerkung: * 0 Subtests bedeutet, dass der Testbogen bzw. die Anmeldung am PC zwar ausgefüllt, aber kein Subtest bearbeitet wurde.

Die Stichprobenstruktur wird in Tabelle 10.2 dargestellt. Deutlich wird, dass der Schülerinnenanteil sehr gering ist. Dies liegt an den Fächerschwerpunkten der Berufsschule (Metall- und Elektrotechnik). Die getesteten Klassen werden in den in Tabelle 10.3 zusammengestellten Berufen ausgebildet.

Tabelle 10.2: Stichprobengröße und Altersstruktur der Berufsschülerstichprobe (Untersuchung 2)

	PBC - CV			CV - PBV		
	weiblich	männlich	unbek.	weiblich	männlich	unbek.
n	18	98	10	15	136	21
	126			172		
	298					
M (SD)	19;1 (1;10)	19;8 (3;6)	19;8 (1;10)	19;3 (2;4)	18;3 (1;9)	18;4 (1;1)
	19;7 (3;2)			18;4 (1;9)		
	18;11 (2;6)					

Bemerkung: Altersangaben in der Form Jahre;Monate

⁷ An dieser Stelle möchte mich bei den Schülern, die an der Untersuchung teilgenommen haben, den Lehrern, die mir ihre Zeit liehen, und insbesondere bei OSTD Siegfried Ullmann bedanken, der mich bei meiner Untersuchung in seiner Schule uneingeschränkt unterstützte.

Die Durchführung der CV in den Klassen hing zum T1 von den Zugangsmöglichkeiten zu den Computerräumen der Schule ab. Die Auswahl der Klassen erfolgte auf Grundlage der Möglichkeiten, Lehrer zu finden, die zeitliche Ressourcen für die Testung erübrigen konnten. Insgesamt sollten die Auswahl und die Zuordnung der Versuchspersonen zu den Bedingungen also als zufällig erachtet werden können. Die Klasse MFM10A wurde hinsichtlich des Modus geteilt, da zum Testzeitpunkt nur ein Computerraum für 15 Schüler zur Verfügung stand und der Zugang zu einem weiteren Raum entgegen der Planung nicht möglich war. Die Testungen fanden in den Wochen vor und nach den Pfingstferien 2007 statt, so dass eine ausreichend große Zeitspanne zwischen den beiden Testungen liegt ($M = 38d$, $SD = 13.6d$; $M(/MME10A) = 42d$, $SD(/MME10A) = 8.4d$). Die Klasse MME10A wurde aufgrund eines Missverständnisses im Abstand von zwei Tagen getestet (vgl. Tabelle 10.3). Dieses Setting soll zur Überprüfung eines möglichen Gedächtniseffektes später genauer betrachtet werden. Insgesamt begannen $n = 172$ Schüler zu T1 mit der CV, $n = 126$ Schüler mit der PBV.

Tabelle 10.3: Ausbildungsrichtung, Klassen, Testmodi, Delay und Klassengrößen der Berufsschülerstichprobe ($n = 298$)

Ausbildung	Klasse	T1	T2	$\Delta(T1,T2)$	n	n_g	n_k	n_{kg}
Elektroniker/in	EGE10C	CV	PBV	33	28	95	23	69
	NGE10C	CV	PBV	34	30		13	
	EBT12B	PBV	CV	49	19		18	
	NEG12	PBV	CV	49	18		15	
Fertigungsmechaniker/in	MFM10A1	CV	PBV	46	15	25	12	18
	MFM10A2	PBV	CV	46	10		6	
Informations- und Telekommunikationssystemelektroniker/in	NIT11	PBV	CV	58	29	29	19	19
Mechatroniker/in	MME10A	PBV	CV	2	25	102	18	66
	MME10B	CV	PBV	47	23		9	
	MME11A	CV	PBV	33	27		21	
	MME11B	CV	PBV	35	27		18	
Mikrotechnolog(e)/in	NMI10	PBV	CV	36	25	47	22	33
	NMI11	CV	PBV	35	22		11	

Bemerkung: n_k – Anzahl der Probanden, die beide Modi vollständig bearbeitet haben ($\sum n_k = 205$)

Es ist klar, dass nicht jeder Proband jeden Subtest in jeder Bedingung der CV bearbeiten kann. Innerhalb der in Kapitel 5.1 und 7.1 beschriebenen Bedingungen wurden die elf Subtests des LPS-neu wie in Tabelle 10.4 dargestellt systematisch variiert. Mit Ausnahme der Subtests 9 und 10 werden so alle Subtests in jeder Bedingung der CV durchgeführt. Für die Stichprobe der Berufsschüler wurde nicht wie bei den Studenten die Reihenfolge der Bedingungen sondern die Reihenfolge der Subtests konstant gehalten. Die Darstellungen und Begrifflichkeiten von Tabelle 10.4 sollen noch kurz an Beispielen erläutert werden: Gruppe 1 bearbeitete Subtest 1 in Bedingung 1: es wurde also jedes Item einzeln präsentiert, durch Anklicken der Lösung erschien das nächste Item und es waren keine Zeitvorgaben ersichtlich. Analog bearbeitete beispielsweise Gruppe 9 Subtest 4 in Bedingung 12: alle Items wurden in eine Kolonne präsentiert, die Lösung blieb ersichtlich, konnte aber nicht mehr verändert werden und die Gesamtzeit für den Subtest wurde angezeigt. Es ist darauf zu achten, dass mit „Bedingung“ immer die Konstellation von Präsentationsform, Antwortformat und Zeitvorgabe gemeint ist. „Gruppe“ bezeichnet dementsprechend das Bedingungsgefüge der elf Subtests.

Tabelle 10.4: Zuweisungsschlüssel der Subtests des LPS-neu zu den Bedingungen der Computerversion

Präsentation		Einzel									Alle					
Antwortformat		Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar	
Zeitvorgabe		Ohne	Gesamt	Item	Ohne	Gesamt	Item	Ohne	Gesamt	Item	Ohne	Gesamt	Ohne	Gesamt	Ohne	Gesamt
Bedingung		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Gruppe	1	1	2	3	4	5	6	7	8	-	-	-	9	10	-	11
	2	11	1	2	3	4	5	6	7	8	-	-	-	9	10	-
	3	-	11	1	2	3	4	5	6	7	8	-	-	-	9	10
	4	9	10	11	1	2	3	4	5	6	7	8	-	-	-	-
	5	-	-	-	11	1	2	3	4	5	6	7	8	9	-	10
	6	-	-	-	10	11	1	2	3	4	5	6	7	8	9	-
	7	10	-	-	-	-	11	1	2	3	4	5	6	7	8	9
	8	-	9	-	-	10	-	11	1	2	3	4	5	6	7	8
	9	8	-	-	-	9	-	10	11	1	2	3	4	5	6	7
	10	7	8	-	-	-	-	9	10	11	1	2	3	4	5	6
	11	6	7	8	9	10	-	-	-	-	11	1	2	3	4	5
	12	5	6	7	8	9	-	10	-	-	-	11	1	2	3	4
	13	4	5	6	7	8	-	9	10	-	-	-	11	1	2	3
	14	3	4	5	6	7	8	-	9	-	-	-	10	11	1	2
	15	2	3	4	5	6	7	8	-	-	-	-	9	10	11	1

Da sich Unterschiede hinsichtlich der Gruppen auf die Effekte der Bedingungen niederschlagen könnten, die trotz der zufälligen Zuordnung zu den Gruppen aufgrund der relativ geringen Fallzahl pro Zelle auftreten können (vgl. Tabelle 10.5), werden die betrachteten Parameter soweit möglich sowohl nach den Bedingungen als auch den Gruppen sortiert zu präsentieren sein. Mittels des a_d -Koeffizienten (Kreuzpointner, Simon & Theis, 2009) kann überprüft werden, ob die unterschiedlichen Gruppengrößen zufällig oder systematisch um den Mittelwert variieren.

Tabelle 10.5: Gruppengrößen der 15 Test-Bedingung-Konstellationen der CV unterteilt nach Testzeitpunkt der jeweiligen Modi bzgl. aller vollständigen Datensätze ($n = 205$)

Gruppe	PBV zu T1	%	PBV zu T2	%
1	7	7.14	9	8.41
2	8	8.16	6	5.61
3	7	7.14	9	8.41
4	7	7.14	5	4.67
5	8	8.16	5	4.67
6	7	7.14	8	7.48
7	8	8.16	10	9.35
8	5	5.10	8	7.48
9	7	7.14	5	4.67
10	7	7.14	7	6.54
11	6	6.12	7	6.54
12	4	4.08	7	6.54
13	5	5.10	8	7.48
14	7	7.14	6	5.61
15	5	5.10	7	6.54
Gesamt	98	100.00	107	100.00

Es ergibt sich ein Wert von $\alpha_d = .974$; der kritische Wert, unter dem eine systematische Variation angenommen werden sollte, beträgt .966. Somit können die Abweichungen als relativ unkritisch erachtet werden.

10.2 Psychometrische Güte der Papier-Bleistift-Version

Bezüglich der Analysen der psychometrischen Eigenschaften der Papier-Bleistift-Version (PBV) werden insbesondere die beiden Bedingungen „PBV zu T1“ und „PBV zu T2“ unterschieden. Im Idealfall ergeben sich keine Unterschiede hinsichtlich der untersuchten Merkmale. Eine erste Voraussetzung für psychometrische Güte ist die Objektivität (z. B. Lienert & Raatz, 1998). Ihr Vorliegen kann im gegebenen Setting nur argumentativ gerechtfertigt werden. Die Durchführungsobjektivität wurde durch eine schriftliche Form der Instruktion, gleiches Testmaterial für alle Probanden, die Aufsicht während der Testung durch den Versuchsleiter und einer Lehrkraft sowie das annähernd gleiche Setting im jeweiligen Klassenzimmer der Berufsschulklassen herzustellen versucht. Die Auswertung erfolgte durch Schablonen und die Eingabe der Lösung der einzelnen Items als richtig oder falsch in eine Datenmatrix. Da in den weiteren Analysen die Rohwerte betrachtet werden, erübrigt sich eine Diskussion der Interpretationsobjektivität, so dass die wichtigsten Merkmale des Gütekriteriums Objektivität als gegeben erachtet werden können.

Die Bestimmung der Reliabilität durch Paralleltests ist bei der alleinigen Betrachtung der PBV nicht möglich, da keine echte parallele Version genutzt wurde. Durch Halbierung der Subtests in gerade und ungerade nummerierte Items können zwei pseudoparallele Versionen geschaffen werden, deren Parallelität (gleiche Mittelwerte und Varianzen) noch zu prüfen sein wird. Die Bestimmung der Retestreliabilität wird erst bei der späteren Berücksichtigung der Computer-version (CV) unter Umständen möglich, falls sich die beiden Modi im Idealfall als parallel erweisen und somit Äquivalenz vorliegt. Als Untergrenze der Reliabilität wird der Koeffizient α (Cronbach, 1951; vgl. Formel (8-18)) bestimmt.

273 der 279 bearbeiteten Testbögen waren vollständig ausgefüllt. Sechs Teilnehmer verspäteten sich, so dass der erste Subtest fehlte. Es werden für die folgenden Analysen nur die $n = 273$ vollständig bearbeiteten Tests genutzt. Tabelle 10.6 gibt einen ersten Überblick über die deskriptive Statistik.

Tabelle 10.6: Deskriptive Statistik der elf Subtests des LPS-neu in der gesamten Berufsschülerstichprobe ($n = 273$) und der Aufteilung in PBV zu T1 ($n = 114$) und PBV zu T2 ($n = 159$)

	Gesamt ($n = 273$)				PBV zu T1 ($n = 114$)				PBV zu T2 ($n = 159$)				$M(\Delta_{T1-T2})$
	Min	Max	M	SD	Min_1	Max_1	M_1	SD_1	Min_2	Max_2	M_2	SD_2	
Subtest 1	7	60	28.33	9.80	7	58	28.12	9.64	9	60	28.48	9.95	-0.36
Subtest 2	4	40	22.78	7.57	7	38	22.49	7.35	4	40	22.99	7.74	-0.50
Subtest 3	10	38	27.01	3.73	10	32	25.82	3.10	17	38	27.86	3.91	-2.03
Subtest 4	8	25	16.68	2.84	11	22	16.43	2.34	8	25	16.86	3.14	-0.43
Subtest 5	12	30	19.06	3.28	12	27	18.86	2.80	12	30	19.21	3.59	-0.35
Subtest 6	2	40	22.43	7.70	2	40	21.99	7.33	5	40	22.74	7.96	-0.75
Subtest 7	11	40	30.57	5.16	13	38	30.02	4.67	11	40	30.97	5.47	-0.95
Subtest 8	4	40	34.43	5.56	15	40	33.29	5.57	4	40	35.25	5.42	-1.96
Subtest 9	0	24	11.23	4.33	0	23	11.21	4.02	0	24	11.25	4.56	-0.04
Subtest 10	3	60	33.15	7.22	3	53	32.34	6.93	11	60	33.72	7.39	-1.38
Subtest 11	0	44	15.99	6.96	0	37	16.25	6.83	0	44	15.81	7.07	0.43

Bemerkungen: ■ - Signifikanzniveau $p < .01$, ■ - Signifikanzniveau $p < .05$, jeweils hinsichtlich des Paarvergleichs der Mittelwerte mittels z-Test (vgl. Tabelle 10.7)

In Tabelle 10.7 werden die inferenzstatistischen Kennwerte zum Vergleich der beiden Stichprobenteile aufgeführt. Lediglich die Subtests 3 und 8 weisen statistisch signifikante Mittelwertsunterschiede auf, so dass ein allgemeiner Gedächtniseffekt durch die wiederholte Durchführung zwar als vorhanden allerdings als begrenzt erachtet werden kann (nur in Subtest 11 ist die Differenz nicht negativ). Die Ergebnisse ohne Berücksichtigung der Klasse MME10a, deren Testungen nur zwei Tage auseinander liegen, ändern sich kaum.

Die Verteilungen der Subtests 3 und 8 werden in Abbildung 10.1 veranschaulicht; die Verteilungen aller Subtests finden sich in Anhang 15.3 und 15.4. In der Grafik wird deutlich, dass die wiederholte Testung von Subtest 3 als PBV insgesamt zu einem besseren Ergebnis der Stichprobe führt. Für den signifikanten Mittelwertsunterschied und die extremen Schiefeffekte bei Subtest 8 dürfte ein Deckeneffekt verantwortlich sein. Als Lösung für dieses Problem könnte die Verkürzung der Bearbeitungsdauer von drei auf zwei Minuten dienen. Dies wird in weiteren Untersuchungen zu prüfen sein.

Tabelle 10.7: z-Test und Effektstärke zum Vergleich der PBV zu T1 und zu T2 sowie Schiefe und Kurtosis der elf Subtests des LPS-neu in der gesamten Berufsschülerstichprobe ($n = 273$) und den Aufteilungen in PBV zu T1 ($n = 114$) und PBV zu T2 ($n = 159$)

	PBV zu T1 vs. PBV zu T2			Gesamt ($n = 273$)		PBV zu T1 ($n = 114$)		PBV zu T2 ($n = 159$)	
	<i>z</i>	<i>p</i>	<i>d'</i>	Schiefe	Kurtosis	Schiefe	Kurtosis	Schiefe	Kurtosis
Subtest 1	0.30	.76	0.04	0.34	-0.25	0.40	-0.02	0.30	-0.37
Subtest 2	0.54	.69	0.07	-0.05	-0.66	-0.02	-0.67	-0.07	-0.65
Subtest 3	4.78	<.01	0.58	-0.09	1.42	-1.20	4.98	0.05	-0.03
Subtest 4	1.30	.34	0.16	-0.20	0.59	-0.20	-0.51	-0.27	0.62
Subtest 5	0.90	.53	0.11	0.34	0.17	0.06	0.05	0.38	-0.03
Subtest 6	0.81	.58	0.10	0.13	-0.37	0.20	-0.02	0.08	-0.54
Subtest 7	1.54	.24	0.19	-0.95	1.40	-0.63	0.83	-1.16	1.80
Subtest 8	2.90	.01	0.36	-1.55	3.64	-1.01	0.45	-2.08	7.34
Subtest 9	0.08	.80	0.01	0.07	0.00	0.08	0.52	0.06	-0.25
Subtest 10	1.58	.23	0.19	-0.03	2.51	-0.68	3.08	0.32	2.05
Subtest 11	0.51	.70	0.06	0.51	1.38	0.26	1.14	0.67	1.64

Bemerkungen: *z* - Statistik des z-Tests auf Gleichheit der Mittelwerte;

d' - Effektstärke nach Cohen (1988): $d' = |M_1 - M_2|/SD'$ mit $SD' = \sqrt{(SD_1^2 + SD_2^2)/2}$;

■ - Signifikanzniveau $\alpha < .01$, ■ - Signifikanzniveau $\alpha < .05$,

Standardfehler der Schiefe: $SE(Gesamt) = .15$, $SE(T1) = .23$, $SE(T2) = .19$;

Standardfehler der Kurtosis: $SE(Gesamt) = .29$, $SE(T1) = .45$, $SE(T2) = .38$;

Für die meisten Subtests liegt zu den beiden Testzeitpunkten Symmetrie, also ein geringer Schiefeparameter vor. Bei Subtest 2 ist der Exzess auffällig. Trotz Symmetrie ist die Verteilung platykurtisch, d. h. breiter verteilt als eine Normalverteilung. Analoges gilt für Subtest 4 zu T1 und Subtest 6 zu T2. Die Verteilungen der Subtests 9 und 11 zu T1 und des Subtests 4 zu T2 sind hingegen leptokurtisch; die Leistung vieler Versuchspersonen liegt somit gehäuft um den Mittelwert bei einigen Ausreißern in beide Richtungen. Die Verteilung von Subtest 1 ist leicht rechtssteil, wobei hierfür wohl eher eine leichte Tendenz zur Bimodalität ursächlich erscheint (vgl. Anhang 15.3). Die linkssteilen Verteilungen von Subtest 7 und Subtest 3 in T1 resultieren vermutlich am ehesten daher, dass es einige Versuchspersonen gab, die sehr gut abschnitten, ein „Totalversagen“ andererseits aufgrund der ersten sehr leichten Aufgaben höchst unwahrscheinlich ist und daher nicht vorkam. Die rechtssteilen Verteilungen der Subtests 5 und 11 zu T2 könnten durch die Zeitbegrenzung bedingt sein, die eine geringere Differenzierung und damit den stärkeren Abfall rechts vom Mittelwert zur Folge haben kann.

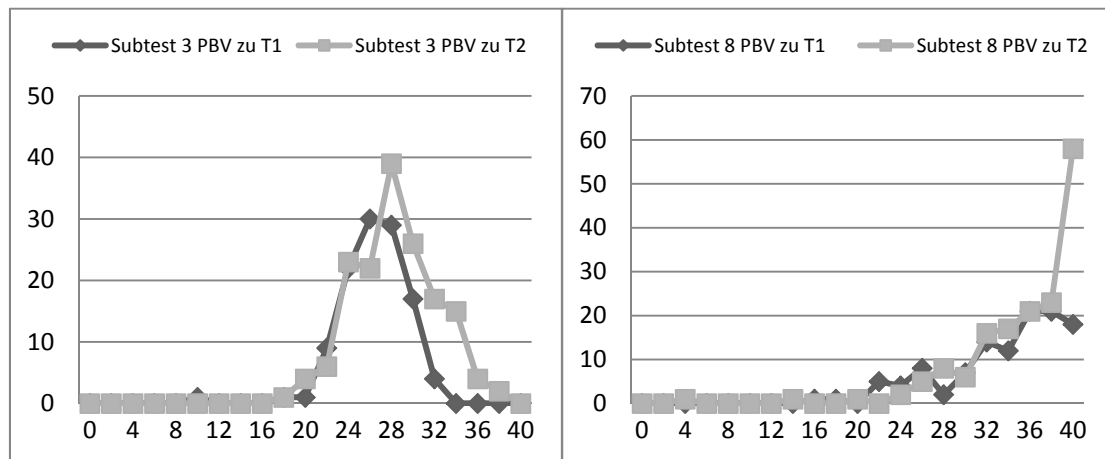


Abbildung 10.1: Verteilungen der Ergebnisse der Subtests 3 und 8 in der Berufsschul-SP unterteilt in PBV zu T1 und PBV zu T2

Auf die Höhe der Kennwerte der Reliabilität haben die Unterschiede der beiden Gruppen PBV zu T1 und PBV zu T2 nur teilweise Einfluss (Subtest 3, PBV zu T1: $\alpha = .67$, $r_{u-g} = .59$, $M_u = 13.01$, $M_g = 12.82$; PBV zu T2: $\alpha = .78$, $r_{u-g} = .71$, $M_u = 13.92$, $M_g = 13.94$; Subtest 8, PBV zu T1: $\alpha = .90$, $r_{u-g} = .86$, $M_u = 16.84$, $M_g = 16.45$; PBV zu T2: $\alpha = .91$, $r_{u-g} = .86$, $M_u = 17.69$, $M_g = 17.57$), insbesondere führt die gemeinsame Betrachtung nur in einem Fall (Subtest 3, PBV zu T1) zu einer Überschätzung (vgl. Tabelle 10.8). Daher werden die Reliabilitätskennwerte der PBV für die Gesamtstichprobe betrachtet. Außer für die Subtests 3, 4 und 5 liegen die Reliabilitätskennwerte im guten bis sehr guten Bereich. Für die Subtests 2, 5 und 6 ist durch die Halbierung in gerade und ungerade Items die Parallelität jedoch nicht ganz angemessen ($d' > 0.2$). Die Reliabilität der Subtests 3, 4 und 5 bewegt sich in einem für die Endfassung eines Leistungstests nicht angemessenen Bereich von .58 bis .76; die nach der Spearman-Brown-Formel korrigierten Split-Half-Indizes geben jedoch Anlass zu Hoffnung, dass dieser psychometrische Mangel in der Endfassung korrigiert werden kann.

Tabelle 10.8: Cronbach's α , Split-Half-Reliabilität und Überprüfung der Parallelität der Teilung der Subtests in die Hälften ungerade (u) und gerade (g) nummerierte Items ($n = 273$)

	α	r_{u-g}	r_{korr}	M_u	SD_u	M_g	SD_g	$M_{\Delta(u-g)}$	$SD_{\Delta(u-g)}$	d'	N_{Items}
Subtest 1	.92	.89	.94	14.17	4.77	14.16	5.31	0.01	5.05	0.00	30
Subtest 2	.90	.87	.93	11.85	3.81	10.93	4.02	0.92	3.92	0.23	20
Subtest 3	.76	.69	.82	13.54	1.97	13.47	2.08	0.07	2.03	0.03	20
Subtest 4	.65	.58	.74	8.48	1.54	8.21	1.65	0.27	1.60	0.17	20
Subtest 5	.74	.71	.83	9.92	1.70	9.15	1.84	0.77	1.77	0.43	20
Subtest 6	.93	.94	.97	11.76	3.88	10.67	3.93	1.09	3.91	0.28	20
Subtest 7	.86	.81	.90	15.18	2.63	15.40	2.79	-0.22	2.71	-0.08	20
Subtest 8	.90	.86	.93	17.33	2.87	17.10	2.89	0.23	2.88	0.08	20
Subtest 9	.80	.89	.94	17.10	2.89	5.65	2.27	0.06	2.60	0.02	20
Subtest 10	.93	.97	.99	16.78	3.62	16.37	3.65	0.41	3.63	0.11	30
Subtest 11	.91	.92	.96	8.23	3.58	7.76	3.52	0.47	3.55	0.13	40

Bemerkung: r_{korr} - nach der Spearman-Brown-Formel ($r_{korr} = 2r_{u-g}/(1 + r_{u-g})$, Lienert & Raatz, 1998) korrigierte Halbierungsreliabilität

Die Itemindizes Trennschärfe und Schwierigkeit der einzelnen Items (vgl. Kapitel 8.6 und Formel (8-20)) werden ihres Umfangs wegen in Anhang 15.5 aufgeführt. Tabelle 10.9 stellt eine Zusammenfassung dar. Bei den Subtests 3, 7 und 8 sind die Aufgaben im Mittel deutlich zu leicht. Für Subtest 3 sind insbesondere die ersten elf Items von fast allen Versuchspersonen gelöst worden. Im 75 %-Quartil kommen offensichtlich nur Flüchtigkeitsfehler vor. Ähnliches gilt für die Subtests 5, 7, 8 und 10. Dort weisen jedoch mehr Items einzelne Falschlösungen auf. Die Zahl der 100 % richtig gelösten Items ist in diesen Subtests deutlich geringer. Die 25 %-Quartile der Subtests 4 und 5 zeigen, dass die Anzahl zu schwieriger Items deutlich zu groß ist. Für Subtest 8 ist sie hingegen zu klein. Die niedrigen Werte der 25 %-Quartile der Subtests 9, 10 und 11 resultieren hingegen eher aus der Besonderheit der Aufgabenstellung. In diesen drei Subtests ist es insbesondere so, dass die Speedkomponente stärker repräsentiert ist und somit die Anzahl der bearbeiteten Items eine gewichtigere Rolle spielt. Somit ist hier die Zahl der Versuchspersonen, die die Items des letzten Drittels bearbeiten, ohnehin geringer und somit der Wert der Schwierigkeit der Items kleiner. Die Korrelation der Positionen der Items mit den Rängen aufgrund der Schwierigkeit für Subtest 1 ist mit $r = .73$ zu niedrig. Für diesen Subtest werden umfangreiche Umstellungen und evtl. der Austausch einiger Items notwendig sein. Die Werte der Korrelationen der übrigen Subtests deuten, wenn überhaupt, nur bedingt auf die Notwendigkeit eines Eingriffs in die Reihenfolge hin.

Tabelle 10.9: Lageparameter der Schwierigkeitskennwerte (p) und Trennschärfen (r_{it}) der elf Subtests des LPS-neu bzgl. der Berufsschul-Stichprobe ($n = 273$)

	Med_p	M_p	SD_p	Min_p	Max_p	$Q_p^{25\%}$	$Q_p^{75\%}$	r_{o-p}	Med_r	$M_{r(it)}$	$SD_{r(it)}$	Min_r	Max_r	$Q_{r(it)}^{25\%}$	$Q_{r(it)}^{75\%}$
Subtest 1	.44	.47	.33	.04	.99	.17	.71	.73	.39	.40	.24	.02	.72	.21	.54
Subtest 2	.61	.57	.27	.07	.98	.38	.81	.91	.36	.41	.20	.02	.66	.31	.54
Subtest 3	.82	.68	.36	.03	1.00	.29	.99	.95	.21	.22	.16	-.08	.52	.12	.29
Subtest 4	.36	.48	.42	.00	1.00	.04	.97	.97	.13	.14	.15	-.23	.44	.01	.25
Subtest 5	.36	.48	.42	.00	1.00	.04	.97	.97	.16	.18	.16	-.07	.41	.02	.34
Subtest 6	.61	.56	.34	.07	.99	.21	.87	.98	.51	.48	.25	.00	.76	.33	.58
Subtest 7	.89	.76	.27	.11	.99	.64	.96	.94	.33	.33	.14	.01	.50	.23	.45
Subtest 8	.92	.86	.16	.39	1.00	.81	.98	.94	.42	.41	.21	-.05	.66	.30	.53
Subtest 9	.14	.28	.32	.00	.94	.01	.52	.95	.23	.22	.20	-.10	.51	.01	.39
Subtest 10	.71	.55	.42	.02	1.00	.05	.97	.98	.43	.42	.21	-.06	.69	.32	.52
Subtest 11	.19	.35	.34	.00	.89	.03	.76	.98	.42	.44	.18	.00	.68	.36	.54

In den Subtests 3, 4 und 5 sind zu viele Items zu schwer und zu leicht; da sich extreme Werte der Itemschwierigkeit negativ hinsichtlich der Trennschärfe auswirken (vgl. Formel (8-20)), weisen diese Subtests entsprechend sowohl im Mittel als auch in den beiden äußeren Quartilsgrenzen geringe Werte auf. Die negativen Werte treten in den Subtests 4 und 5 bei Items auf, die jenseits der Grenze liegen, innerhalb derer die Probanden die Items wirklich bearbeitet hatten, sind also insofern unkritisch und durch Rateeffekte zu erklären. Der negative Wert in Subtest 8 resultiert aus der einzigen falschen Lösung des vierten Items durch eine Versuchsperson, deren Subtestwert über dem Mittelwert liegt. Dies kann als Flüchtigkeitsfehler und als solches ebenso unkritisch erachtet werden. Für Subtest 10 liegt der einzige negative Wert für Item 7 vor, bei dem von sieben Probanden eine falsche Ziffer nicht markiert wurde. Ein spezieller Grund hierfür ist jedoch nicht ersichtlich. Die Trennschärfen für Subtest 10 sind aufgrund des Speedcharakters der

Tests und dessen linearen Aufbaus allgemein recht hoch. Für Subtest 9 sind die Werte insbesondere für die Items gut, bei denen Einser, Zweier, Dreier und Vierer zu zählen waren. Die höheren Ziffern wurden nur sehr selten erreicht, die negativen Trennschärfewerte ab Item 33 resultieren wohl aufgrund von Raten, da sie von nur mehr höchstens zwei Versuchspersonen angekreuzt wurden. Die Trennschärfen der übrigen Subtests liegen insbesondere in Anbetracht der Speedkomponente der Aufgaben in angemessenen Quartilsgrenzen, wobei eine Verschiebung insbesondere des 75 %-Quartils von Vorteil wäre, so dass die Zahl der sehr leichten Items nicht zu hoch ist.

Insgesamt haben sich sowohl die übernommenen (3, 6, 7, 8, 9 und 10), als auch die veränderten Subtests (1, 2 und 11) psychometrisch bewährt. In den Subtests 4 und 5 müssen vor einer Normierung noch Änderungen an den zu schweren Items vorgenommen werden, damit sich die Reliabilität an das Niveau der übrigen Subtests und somit an die geforderte Höhe bewegt. Die zweite Testung als PBV führt im Mittel nur zu geringfügig höheren Rohwerten. Für Subtest 8 könnte die Verkürzung der Bearbeitungszeit sowohl die Unterschiede zweier Durchführungen verringern, als auch die Verteilung der Leistungen über die Stichprobe verbessern. Subtest 3 zeigt bei der Testung als PBV zu T1 eine vermeintlich passende Verteilung (vgl. Abbildung 10.1). Durch einige Werte im unteren Bereich und das Maximum von 32 Rohwertpunkten liegen die Ausprägungen von Schiefe und Kurtosis jedoch deutlich über den geeigneten Ausprägungen ($sk = -1.20$, $kurt = 4.98$). Die platykurtische rechtssteile Verteilung zu T1 im Vergleich zur relativ symmetrischen Verteilung zu T2, aufgrund dessen der p -Value verteilungsbedingt niedriger ausfällt, lässt jedoch hoffen, dass es sich in diesem Kontext um eine stichprobenzufällige Abweichung handelt, die in weiteren Untersuchungen und bei größeren Stichproben nicht mehr vorkommen wird. Mit Ausnahme von Subtest 8 als PBV zu T2 scheinen alle Subtests zu den beiden Messzeitpunkten einigermaßen angemessen verteilt zu sein (vgl. Anhang 13.4). Dennoch zeigen einige Subtests (insbesondere 7, 10 und 11) auffällige Verteilungsparameter (vgl. Tabelle 10.7), die auch bei der gemeinsamen Betrachtung (und somit der Erhöhung der Fallzahl) nicht gänzlich verschwinden. Für Subtest 7 wäre eine Verkürzung der Bearbeitungszeit eine mögliche Lösung, da auch hier die Tendenz zu einem Deckeneffekt auszumachen ist. Der leptokurtische Verlauf von Subtest 10 wird wohl aufgrund der Testkonstruktion hingenommen werden müssen. Veränderungen in der Bearbeitungszeit hätten bei einer Verkürzung eine noch stärkere Bündelung um den Mittelwert und bei einer Verlängerung vermutlich eine einfache Verschiebung mit dem Risiko eines Deckeneffekts zur Folge. Für Subtest 11 hätte eine Verlängerung der Bearbeitungszeit voraussichtlich positive Effekte hinsichtlich der Verteilung. Es würden mehr Werte größer als der Mittelwert resultieren, die die Schiefe und im Idealfall die Kurtosis verringern.

Generell scheint der Anteil der Speedkomponente mit der Verzerrung der Verteilung weg von symmetrisch und normal einherzugehen. Dennoch zeigt sich, dass die Subtests über zwei Testungen hinweg vergleichbare Ergebnisse erwarten lassen. Inwiefern dies in der Testung als CV der Fall ist, wird im Folgenden dargestellt.

10.3 Psychometrische Güte der Computer-Version

Aufgrund der unterschiedlichen Durchführungsbedingungen ist eine zusammenfassende Interpretation der Subtestergebnisse der CV hinsichtlich ihrer Reliabilität erst dann zulässig, wenn nachgewiesen ist, dass sich die Resultate der Subtests in den verschiedenen Bedingungen nicht unterscheiden. Auch die Objektivität ist in gewisser Weise durch die Variation der Darstellung betroffen. Hinsichtlich der Durchführungsobjektivität können die notwendigen Unterschiede der Beschreibungen der Bedingungen Einfluss auf die Lösungsstrategien genommen haben. Dieser Einfluss ist jedoch mit den Bedingungen verschränkt und kann durch das gewählte Setting nicht kontrolliert werden. Da das vornehmliche Ziel jedoch die Äquivalenz mit der PBV ist, ist es gerechtfertigt, die Objektivität der einzelnen Bedingungen zu betrachten. Diese kann für alle drei Aspekte als gegeben erachtet werden. Alle Probanden hatten hinsichtlich der Durchführung im selben Setting dieselbe Instruktion, die am Computer vorgegeben wurde. Die Ergebnisse wurden durch den PC aufgezeichnet und ausgewertet, wodurch eine zufällige Falschkodierung ausgeschlossen ist. Die Rohwertsumme als zu vergleichender Messwert ist hinsichtlich der Interpretation unkritisch.

Wegen der geringen Fallzahlen pro Subtest pro Bedingung ist die messfehlerunkritische Bestimmung der Reliabilität und der Itemkennwerte erschwert. Für jede der 15 Bedingungen stehen zwischen elf und 18 Datensätze zur Verfügung (vgl. Tabelle 10.5). Die Betrachtung der Kennwerte der einzelnen Items erscheint aufgrund der zu erwartenden Messfehler als wenig aussagekräftig. Die Mittelwerte der Schwierigkeits- und Trennschärfeparameter können hingegen einen ersten deskriptiven Überblick über die Ähnlichkeit der Durchführungsbedingungen der CV ebenso wie hinsichtlich der PBV geben.

Tabelle 10.10: Cronbach's α der Subtests des LPS-neu in den 15 Vpn-Gruppen (definiert durch die Bedingungskombinationen in den einzelnen Subtests) der CV

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	PBV
Subtest 1	.97	.90	.95	.95	.93	.92	.94	.94	.94	.95	.94	.92	.96	.94	.93	.92
Subtest 2	.92	.89	.92	.93	.89	.93	.92	.94	.86	.94	.94	.88	.93	.95	.90	.90
Subtest 3	.88	.53	.75	.80	.74	.73	.78	.75	.54	.84	.73	.70	.72	.64	.73	.76
Subtest 4	.79	.57	.66	.70	.80	.48	.61	.41	.53	.76	.63	.62	.69	.78	.38	.65
Subtest 5	.73	.85	.84	.78	.78	.52	.75	.61	.58	.83	.77	.61	.43	.42	.73	.74
Subtest 6	.91	.87	.94	.89	.91	.92	.93	.84	.95	.93	.91	.93	.87	.96	.92	.93
Subtest 7	.81	.74	.78	.81	.89	.76	.85	.91	.57	.84	.90	.86	.82	.87	.67	.86
Subtest 8	.89	.83	.85	.84	.94	.96	.92	.95	.82	.90	.87	.72	.86	.88	.81	.90
Subtest 9	.47	.64	.56	.59	.77	.73	.68	.93	.81	.86	.87	.89	.87	.89	.47	.80
Subtest 10	.91	.94	.92	.89	.94	.90	.91	.90	.98	.97	.93	.98	.98	.92	.92	.93
Subtest 11	.93	.87	.95	.95	.91	.91	.90	.91	.93	.91	.89	.85	.91	.92	.89	.91

Bemerkungen: Die Berechnung berücksichtigt nicht die Reihenfolge der Testmodi (PBV-CV bzw. CV-PBV).
 $\alpha < .70$ grau markiert

Auch für die Betrachtung der internen Konsistenz darf die geringe Stichprobengröße zwischen 16 und 20 Versuchspersonen (ungeachtet des Messzeitpunktes) nicht außer Acht gelassen werden. Keine Gruppe fällt durch konsistent geringere Werte im Vergleich zu den anderen Gruppen

auf (vgl. Tabelle 10.10), so dass die Werte kleiner .70 durch die Interaktion Test und Bedingung verursacht scheinen oder Stichprobeneffekte sind. Inwiefern diese Interaktionen gruppenspezifisch sind, kann durch die Betrachtung von Tabelle 10.11 erschlossen werden. Für Subtest 1 resultieren in jeder Bedingung angemessene Werte, der kleinste Wert ergibt sich für Gruppe 2. Analoges gilt für Subtest 2, wobei hier Bedingung 3 durch Gruppe 2 durchlaufen wurde. Für Subtest 3 zeigen sich große Schwankungen; das Spektrum reicht von .95 bis .53 (Gruppe 2!), während die PBV ein α von .76 aufweist. Ähnliche Ergebnisse stellten sich für die Subtests 4 und 5 ein, es zeigt sich jedoch kein Muster hinsichtlich der Bedingungen. Für die Bedingungen 7 und 8 weist die CV tendenziell höhere Werte als die PBV auf.

Tabelle 10.11: Cronbach's α der Subtests des LPS-neu in den 15 Bedingungen der CV

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	PBV
Subtest 1	.97	.90	.95	.95	.93	.92	.94	.94	.94	.95	.94	.92	.96	.94	.93	.92
Subtest 2	.90	.92	.89	.92	.93	.89	.93	.92	.94	.86	.94	.94	.88	.93	.95	.90
Subtest 3	.64	.73	.88	.53	.75	.80	.74	.73	.78	.75	.54	.84	.73	.70	.72	.76
Subtest 4	.69	.78	.38	.79	.57	.66	.70	.80	.48	.61	.41	.53	.76	.63	.62	.65
Subtest 5	.61	.43	.42	.73	.73	.85	.84	.78	.78	.52	.75	.61	.58	.83	.77	.74
Subtest 6	.91	.93	.87	.96	.92	.91	.87	.94	.89	.91	.92	.93	.84	.95	.93	.93
Subtest 7	.84	.90	.86	.82	.87	.67	.81	.74	.78	.81	.89	.76	.85	.91	.57	.86
Subtest 8	.82	.90	.87	.72	.86	.88	.81	.89	.83	.85	.84	.94	.96	.92	.95	.90
Subtest 9	.59	.93	-	.87	.86	-	.86	.87	-	-	-	.47	.71	.65	.68	.80
Subtest 10	.91	.89	-	.90	.91	-	.98	.98	-	-	-	.92	.92	.94	.92	.93
Subtest 11	.87	.95	.95	.91	.91	.90	.91	.93	.91	.89	.85	.91	.92	.89	.73	.91

Bemerkungen: Die Berechnung berücksichtigt nicht die Reihenfolge der Testmodi (PBV-CV bzw. CV-PBV).

$\alpha < .70$ grau markiert

Auch die Subtests 6, 10 und 11 weisen generell relativ hohe und der PBV entsprechende Werte auf. Für Subtest 9 schwanken die Werte sehr stark. Einige Bedingungen weisen relativ niedrige Werte auf. In vielen Bedingungen resultieren hingegen sehr gute Konsistenzwerte.

Insgesamt findet sich also das Bild der PBV bestätigt, dass sich die bestehenden und leicht veränderten Subtests des LPS-neu bewährt haben. Einschränkungen müssen nur für Subtest 4, 5 und 9 und einige einzelne Bedingungskonstellationen gemacht werden.

Keine der Gruppe fällt dadurch auf, dass die Items aller Subtests für sie besonders schwer oder leicht waren, wobei doch einige Tendenzen ersichtlich sind (vgl. Tabelle 10.12). Die Gruppen 2, 3, 4, 5 und 12 weisen bei der CV zu T1 bei den meisten Subtests mittlere Itemschwierigkeiten über dem Durchschnitt aller Gruppen auf. Bei den Gruppen 2 und 3 ist die Differenz von fünf, bei 5 und 12 von drei und bei 4 von zwei Subtests größer als eine Standardabweichung. Die Gruppen 6, 11 und 14 haben in der CV zu T1 einige Schwierigkeitswerte, die mehr als eine Standardabweichung kleiner als der Mittelwert sind (6 sechs, 11 und 14 je vier). Die übrigen Subtests dieser Gruppen sind jedoch durch- oder teilweise deskriptiv überdurchschnittlich. In der CV zu T2 zeigt sich dieser Effekt nur bei Gruppe 12. Die Werte der Gruppen 2 liegen hier eher unter und im Durchschnitt. Die mittlere Schwierigkeit der Gruppe 3 liegt bei neun der elf Subtests leicht über, bei zwei Subtests deutlich unter dem Durchschnitt. Des Weiteren fallen durch eher niedrige Werte in der CV zu T2 die Gruppen 4, 8 und 13 auf. Die Gruppen 1 und 14 zeigen

hier hingegen je fünf überdurchschnittliche Werte, allerdings auch einige unterdurchschnittliche.

Tabelle 10.12: Mittlere Schwierigkeit der CV und der PBV der Subtests des LPS-neu in den 15 Vpn-Gruppen der CV

		CV															PBV																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	M
CV zu T1, PVB zu T2	1	.26	.31	.41	.28	.30	.30	.29	.32	.26	.33	.31	.43	.39	.33	.31	.32 (.05)	.40	.45	.52	.46	.44	.44	.53	.44	.47	.50	.43	.55	.50	.46	.53	.47 (.04)
	2	.44	.55	.47	.49	.46	.37	.47	.49	.39	.44	.39	.48	.31	.40	.52	.44 (.06)	.60	.73	.69	.61	.65	.45	.55	.55	.60	.57	.53	.78	.51	.51	.63	.60 (.09)
	3	.65	.69	.69	.64	.67	.59	.59	.64	.65	.63	.59	.64	.64	.66	.65	.64 (.03)	.68	.76	.78	.65	.76	.65	.70	.74	.71	.69	.67	.73	.69	.73	.70	.71 (.04)
	4	.40	.42	.46	.41	.42	.37	.41	.40	.43	.41	.36	.42	.40	.41	.40	.41 (.02)	.43	.45	.47	.44	.42	.42	.41	.43	.44	.44	.39	.48	.41	.46	.47	.44 (.03)
	5	.45	.53	.48	.49	.48	.43	.46	.43	.45	.42	.46	.48	.49	.45	.47	.46 (.03)	.47	.56	.54	.51	.56	.47	.44	.47	.52	.48	.48	.52	.48	.46	.54	.50 (.04)
	6	.44	.53	.50	.47	.52	.49	.43	.46	.54	.46	.31	.49	.42	.31	.46	.45 (.07)	.60	.69	.64	.67	.66	.70	.59	.52	.63	.53	.42	.58	.52	.37	.69	.59 (.10)
	7	.68	.71	.74	.71	.77	.69	.74	.68	.83	.70	.60	.74	.66	.75	.73	.71 (.05)	.81	.88	.83	.77	.84	.78	.84	.78	.89	.81	.73	.80	.76	.77	.76	.80 (.05)
	8	.76	.84	.82	.78	.80	.73	.79	.73	.77	.78	.80	.83	.77	.73	.80	.78 (.04)	.90	.93	.88	.88	.91	.86	.93	.93	.86	.89	.85	.94	.87	.88	.87	.89 (.03)
	9	.29	.31	.34	.39	.39	.37	.39	.21	.27	.13	.24	.27	.11	.13	.39	.28 (.10)	.53	.50	.49	.62	.69	.51	.59	.48	.64	.60	.56	.58	.41	.51	.56	.55 (.07)
	10	.27	.25	.28	.32	.33	.38	.46	.40	.44	.39	.39	.45	.40	.26	.27	.35 (.07)	.54	.59	.63	.62	.68	.52	.59	.60	.53	.57	.54	.55	.54	.54	.59	.58 (.04)
	11	.27	.21	.37	.35	.35	.18	.30	.21	.33	.26	.27	.29	.27	.31	.33	.29 (.06)	.39	.48	.51	.43	.52	.37	.36	.39	.45	.44	.45	.44	.37	.45	.53	.44 (.06)
PBV zu T1, CV zu T2	1	.72	.43	.54	.47	.54	.47	.46	.53	.44	.60	.51	.56	.42	.60	.49	.52 (.08)	.62	.40	.50	.41	.47	.45	.52	.45	.44	.52	.45	.53	.45	.51	.50	.48 (.06)
	2	.76	.69	.69	.56	.63	.63	.69	.66	.54	.69	.63	.80	.57	.78	.64	.66 (.08)	.69	.63	.67	.56	.58	.50	.55	.57	.54	.58	.53	.71	.51	.56	.59	.58 (.06)
	3	.66	.71	.71	.69	.73	.70	.70	.65	.68	.69	.74	.71	.74	.73	.73	.70 (.03)	.66	.71	.73	.65	.69	.65	.68	.69	.65	.68	.66	.69	.69	.70	.62	.68 (.03)
	4	.48	.46	.38	.43	.44	.39	.45	.44	.43	.45	.44	.48	.47	.50	.41	.44 (.03)	.43	.44	.43	.43	.41	.43	.41	.41	.41	.44	.40	.45	.42	.44	.43	.43 (.01)
	5	.58	.41	.52	.46	.48	.47	.50	.41	.50	.54	.52	.54	.49	.50	.46	.49 (.05)	.50	.52	.53	.48	.48	.46	.47	.46	.48	.48	.48	.53	.47	.48	.49	.49 (.02)
	6	.59	.57	.60	.49	.56	.68	.60	.50	.59	.57	.53	.61	.43	.61	.66	.57 (.06)	.60	.63	.61	.57	.58	.66	.57	.53	.57	.53	.39	.57	.49	.48	.68	.56 (.07)
	7	.78	.76	.81	.73	.77	.79	.79	.76	.76	.77	.84	.84	.77	.76	.81	.78 (.03)	.76	.80	.83	.75	.76	.77	.82	.75	.81	.76	.76	.81	.75	.75	.75	.78 (.03)
	8	.89	.84	.90	.82	.83	.90	.92	.93	.90	.91	.89	.94	.88	.86	.93	.89 (.04)	.78	.90	.88	.82	.85	.84	.91	.89	.83	.85	.87	.92	.86	.85	.88	.86 (.04)
	9	.36	.38	.35	.38	.37	.36	.35	.31	.29	.23	.37	.34	.22	.24	.37	.33 (.06)	.49	.54	.47	.53	.53	.56	.57	.49	.56	.57	.59	.59	.46	.51	.51	.53 (.04)
	10	.34	.26	.29	.31	.33	.45	.45	.47	.59	.54	.55	.61	.57	.35	.36	.40 (.09)	.57	.51	.59	.57	.59	.55	.55	.57	.54	.51	.54	.55	.55	.56	.60	.56 (.03)
	11	.56	.29	.46	.38	.44	.47	.44	.28	.37	.43	.45	.52	.41	.55	.34	.43 (.08)	.50	.38	.50	.39	.45	.41	.40	.38	.39	.42	.44	.48	.40	.47	.46	.43 (.04)

Bemerkungen: Farblich hinterlegte Werte weichen auffällig vom Mittelwert der entsprechenden PBV ab; CV zu T1 wird mit den Mittelwerten der PBV zu T1 verglichen; entsprechend CV zu T2 mit PBV zu T2;

$\Delta(p_i - M) < -.1$: dunkelgrau; $\Delta(p_i - M) < -.05$: hellgrau;

$\Delta(p_i - M) > .05$: hellorange; $\Delta(p_i - M) > .1$: dunkelorange

Für die Ordnung der mittleren Itemschwierigkeiten nach den Durchführungsbedingungen lassen sich fast keine Häufungen von abweichenden Werten pro Bedingung ausmachen. Lediglich in der CV zu T2 in den Bedingungen 6 und 9 liegen die Werte nur vier bzw. zwei Mal über dem Durchschnitt. Die übrigen mittleren Schwierigkeiten der Subtests sind entweder nur etwas oder in zwei bzw. drei Fällen sogar mehr als eine Standardabweichung kleiner als der Durchschnitt (vgl. Tabelle 10.13).

Es zeigen sich somit so gut wie keine systematischen Bedingungseffekte hinsichtlich der mittleren Schwierigkeiten der Items innerhalb der CV. Für einige Gruppen erweisen sich die Items im Mittel etwas leichter, wobei dies lediglich für Gruppe 12 in beiden Reihenfolgepositionen der Fall ist. Ob dies eine zufällige Kumulierung stärkerer Probanden ist oder dem der Gruppe zugrunde liegenden Bedingungsgefüge geschuldet ist, wird im Rahmen des Vergleichs der Leistungswerte der Subtest diskutiert.

Tabelle 10.13: Mittlere Schwierigkeit der CV und der PBV der Subtests des LPS-neu in den 15 Bedingungen der CV

		CV																PBV															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	M	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	M
CV zu T1, PVB zu T2	1	.26	.31	.41	.28	.30	.30	.29	.32	.26	.33	.31	.43	.39	.33	.31	.32 (.05)	.40	.45	.52	.46	.44	.44	.53	.44	.47	.50	.43	.55	.50	.46	.53	.47 (.04)
	2	.52	.44	.55	.47	.49	.46	.37	.47	.49	.39	.44	.39	.48	.31	.40	.44 (.06)	.63	.60	.73	.69	.61	.65	.45	.55	.55	.60	.57	.53	.78	.51	.51	.60 (.09)
	3	.66	.65	.65	.69	.69	.64	.67	.59	.59	.64	.65	.63	.59	.64	.64	.64 (.03)	.73	.70	.68	.76	.78	.65	.76	.65	.70	.74	.71	.69	.67	.73	.69	.71 (.04)
	4	.40	.41	.40	.40	.42	.46	.41	.42	.37	.41	.40	.43	.41	.36	.42	.41 (.02)	.41	.46	.47	.43	.45	.47	.44	.42	.42	.41	.43	.44	.44	.39	.48	.44 (.03)
	5	.48	.49	.45	.47	.45	.53	.48	.49	.48	.43	.46	.43	.45	.42	.46	.46 (.03)	.52	.48	.46	.54	.47	.56	.54	.51	.56	.47	.44	.47	.52	.48	.48	.50 (.04)
	6	.31	.49	.42	.31	.46	.44	.53	.50	.47	.52	.49	.43	.46	.54	.46	.45 (.07)	.42	.58	.52	.37	.69	.60	.69	.64	.67	.66	.70	.59	.52	.63	.53	.59 (.10)
	7	.70	.60	.74	.66	.75	.73	.68	.71	.74	.71	.77	.69	.74	.68	.83	.71 (.05)	.81	.73	.80	.76	.77	.76	.81	.88	.83	.77	.84	.78	.84	.78	.89	.80 (.05)
	8	.77	.78	.80	.83	.77	.73	.80	.76	.84	.82	.78	.80	.73	.79	.73	.78 (.04)	.86	.89	.85	.94	.87	.88	.87	.90	.93	.88	.88	.91	.86	.93	.93	.89 (.03)
	9	.39	.21	-	.24	.27	-	.12	.13	-	-	-	.34	.35	.36	.39	.28 (.10)	.62	.48	-	.56	.61	-	.51	.51	-	-	-	.55	.60	.50	.59	.55 (.07)
	10	.46	.32	-	.38	.40	-	.45	.40	-	-	-	.26	.27	.25	.28	.35 (.07)	.59	.62	-	.52	.57	-	.54	.56	-	-	-	.54	.57	.59	.63	.58 (.04)
	11	.21	.37	.35	.35	.18	.30	.21	.33	.26	.27	.29	.27	.31	.33	.40	.29 (.06)	.48	.51	.43	.52	.37	.36	.39	.45	.44	.45	.44	.37	.45	.53	.43	.44 (.06)
PBV zu T1, CV zu T2	1	.72	.43	.54	.47	.54	.47	.46	.53	.44	.60	.51	.56	.42	.60	.49	.52 (.08)	.62	.40	.50	.41	.47	.45	.52	.45	.44	.52	.45	.53	.45	.51	.50	.48 (.06)
	2	.64	.76	.69	.69	.56	.63	.63	.69	.66	.54	.69	.63	.80	.57	.78	.66 (.08)	.59	.69	.63	.67	.56	.58	.50	.55	.57	.54	.58	.53	.71	.51	.56	.58 (.06)
	3	.73	.73	.66	.71	.71	.69	.73	.70	.70	.65	.68	.69	.74	.71	.74	.70 (.03)	.70	.62	.66	.71	.73	.65	.69	.65	.68	.69	.65	.68	.66	.69	.69	.68 (.03)
	4	.47	.50	.41	.48	.46	.38	.43	.44	.39	.45	.44	.43	.45	.44	.48	.44 (.03)	.42	.44	.43	.43	.44	.43	.43	.41	.43	.41	.41	.41	.44	.40	.45	.43 (.01)
	5	.54	.49	.50	.46	.58	.41	.52	.46	.48	.47	.50	.41	.50	.54	.52	.49 (.05)	.53	.47	.48	.49	.50	.52	.53	.48	.48	.46	.47	.46	.48	.48	.48	.49 (.02)
	6	.53	.61	.43	.61	.66	.59	.57	.60	.49	.56	.68	.60	.50	.59	.57	.57 (.06)	.39	.57	.49	.48	.68	.60	.63	.61	.57	.58	.66	.57	.53	.57	.53	.56 (.07)
	7	.77	.84	.84	.77	.76	.81	.78	.76	.81	.73	.77	.79	.79	.76	.76	.78 (.03)	.76	.76	.81	.75	.75	.75	.76	.80	.83	.75	.76	.77	.82	.75	.81	.78 (.03)
	8	.90	.91	.89	.94	.88	.86	.93	.89	.84	.90	.82	.83	.90	.92	.93	.89 (.04)	.83	.85	.87	.92	.86	.85	.88	.78	.90	.88	.82	.85	.84	.91	.89	.86 (.04)
	9	.38	.31	-	.37	.32	-	.23	.24	-	-	-	.37	.38	.36	.35	.33 (.06)	.53	.49	-	.59	.58	-	.52	.51	-	-	-	.50	.54	.52	.57	.53 (.04)
	10	.45	.31	-	.45	.51	-	.60	.56	-	-	-	.35	.35	.26	.29	.40 (.09)	.55	.57	-	.55	.56	-	.55	.53	-	-	-	.56	.59	.51	.59	.56 (.03)
	11	.29	.46	.38	.44	.47	.44	.28	.37	.43	.45	.52	.41	.55	.34	.48	.43 (.08)	.38	.50	.39	.45	.41	.40	.38	.39	.42	.44	.48	.40	.47	.46	.43	.43 (.04)

Bemerkungen: Farblich hinterlegte Werte weichen auffällig vom Mittelwert der entsprechenden PBV ab; CV zu T1 wird mit den Mittelwerten der PBV zu T1 verglichen; entsprechend CV zu T2 mit PBV zu T2;

$\Delta(p_i - M) < -.1$: dunkelgrau; $\Delta(p_i - M) < -.05$: hellgrau;

$\Delta(p_i - M) > .05$: hellorange; $\Delta(p_i - M) > .1$: dunkelorange

Bei der Betrachtung der Trennschärfen über alle Bedingungen hinweg resultieren für alle Subtests vorwiegend akzeptable Werte (vgl. Tabelle 10.14). Die negativen Werte der Minima der Subtests 4 und 9 können als unkritisch gesehen werden, da beide Items am Ende der Subtests standen und von nur mehr einigen Versuchspersonen bearbeitet wurden, wobei die Durchsicht der entsprechenden Antwortmuster zeigt, dass hierbei offensichtlich auch geraten wurde. Maxima und 75 %-Quartile zeigen, dass die Trennschärfen bei einer Betrachtung der Subtests über

alle Bedingungen hinweg durchaus hoch sind. Die Berücksichtigung beider Quartilsgrenzen führt zu dem Schluss, dass die Mehrheit der Items bei einer Gesamtbetrachtung eine angemessene Trennschärfe aufweist.

Tabelle 10.14: Lageparameter der Trennschärfen der Subtests des LPS-neu für die CV einzeln berechnet für die 15 Gruppen

		<i>Med_{r(it)}</i>	<i>M_{r(it)}</i>	<i>SD_{r(it)}</i>	<i>Min_{r(it)}</i>	<i>Max_{r(it)}</i>	<i>Q_{r(it)}^{25%}</i>	<i>Q_{r(it)}^{75%}</i>
Subtest	1	.41	.41	.21	.10	.80	.26	.59
	2	.46	.44	.14	.19	.68	.38	.62
	3	.37	.38	.16	.06	.67	.28	.52
	4	.27	.25	.10	-.01	.41	.18	.34
	5	.32	.31	.15	.07	.55	.20	.47
	6	.44	.43	.15	.19	.73	.34	.58
	7	.57	.49	.17	.08	.77	.40	.67
	8	.58	.51	.09	.35	.71	.50	.62
	9	.28	.30	.25	-.06	.66	.08	.47
	10	.68	.59	.15	.23	.84	.55	.82
	11	.50	.42	.15	.00	.63	.40	.55

Bemerkungen: Zur Berechnung der Mittelwerte und Standardabweichungen wurden die Trennschärfeparameter der einzelnen Items mittels Fisher-Transformation umgerechnet (vgl. Kapitel 7.4); die Ergebnisse wurden für die bessere Interpretierbarkeit wieder zurückgerechnet; die Perzentilwerte basieren auf den unveränderten Korrelationen

Werden die Trennschärfeparameter unter Berücksichtigung der unterschiedlichen Durchführungsbedingungen berechnet, fällt zunächst auf, dass sie teilweise deutlich höhere Werte annehmen, allerdings auch deutlich stärker streuen (vgl. Tabelle 10.15 und Tabelle 10.16). Um die Bedeutung der starken Schwankungen und der teils hohen negativen Werte besser einordnen zu können, soll kurz auf die Berechnung der Trennschärfe eingegangen werden (vgl. Formel (8-20)). Die Trennschärfe wird dann negativ, wenn der Gesamtscore des Subtests der Probanden, die das Item nicht gelöst haben, höher ist als der Gesamtscore der Probanden, die das Item gelöst haben. Die Wurzel der Varianz der Itemschwierigkeit ist maximal .50, wenn die Schwierigkeit im Mittel, also .50, ist. Für sehr schwere Items resultiert ebenso wie für sehr leichte Items ein deutlich geringerer Wert. Bei maximal 18 Versuchspersonen in einer Gruppe wäre der Faktor bei einer Falschlösung beispielsweise $\sqrt{17/18(1 - 17/18)} = \sqrt{17/324} \approx .23$, für die kleinste Gruppe mit $n = 12$ liegt der kleinste Faktor bei rd. .28. Je nach Itemschwierigkeit wird die Differenz der Subtestscores somit auf ein Viertel bis zur Hälfte reduziert. Des Weiteren wird dieses Ergebnis durch die Standardabweichung des Subtests geteilt. Je größer die Streuung, desto geringer die Trennschärfe. Durch die Verringerung der Stichprobengröße bei der Betrachtung der Gruppen hinsichtlich der Durchführungsbedingungen resultieren starke Unterschiede in den Streuungen der Items und somit teilweise eine Verringerung und teilweise eine Erhöhung des Betrags der Trennschärfe. Dies kann die deutlich größere Bandbreite der Werte in Tabelle 10.15 und Tabelle 10.16 erklären. Die unterschiedlichen Items können jedoch auch unterschiedlich in den einzelnen Bedingungen reagieren und somit für die jeweiligen Konstellationen unterschiedlich, aber dennoch jeweils gut sein. Bei der gemeinsamen Betrachtung resultiert in der Folge ein niedriger Wert der Trennschärfe.

Tabelle 10.15: Lageparameter der Trennschärfe der Subtests der CV des LPS-neu unterteilt in die 15 Gruppen der CV

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.62 (.44) [-.10;.91] {.36;.76}	.34 (.37) [-.53;.79] {.00;.53}	.51 (.42) [-.33;.86] {.27;.74}	.49 (.40) [-.10;.84] {.10;.65}	.38 (.38) [-.38;.83] {.00;.66}	.36 (.37) [-.15;.80] {.00;.57}	.47 (.48) [-.07;.89] {.00;.75}	.47 (.41) [-.23;.85] {.00;.70}	.52 (.32) [-.10;.74] {.39;.68}	.46 (.46) [-.45;.86] {.00;.71}	.41 (.41) [-.38;.84] {.00;.60}	.42 (.43) [-.40;.87] {.00;.70}	.48 (.34) [-.14;.75] {.22;.70}	.45 (.39) [-.15;.82] {.00;.66}	.44 (.32) [-.07;.79] {.28;.60}
2	.51 (.45) [-.36;.88] {.26;.73}	.43 (.28) [-.09;.71] {.28;.62}	.48 (.33) [-.05;.83] {.29;.63}	.52 (.38) [-.03;.84] {.30;.78}	.46 (.32) [-.19;.71] {.19;.65}	.53 (.33) [.00;.82] {.33;.74}	.51 (.37) [-.09;.85] {.26;.68}	.53 (.28) [-.18;.79] {.38;.65}	.35 (.36) [-.44;.71] {.00;.62}	.53 (.29) [.00;.80] {.36;.66}	.55 (.25) [.00;.76] {.44;.62}	.43 (.41) [-.25;.87] {.00;.62}	.48 (.31) [-.38;.79] {.33;.62}	.61 (.30) [.00;.85] {.45;.73}	.39 (.38) [-.68;.78] {.03;.62}
3	.46 (.38) [-.39;.83] {.23;.66}	.18 (.35) [-.34;.84] {.00;.34}	.27 (.31) [-.07;.72] {.00;.45}	.23 (.32) [-.16;.81] {.00;.43}	.16 (.24) [-.11;.69] {.00;.28}	.24 (.27) [-.09;.69] {.00;.44}	.27 (.28) [-.07;.79] {.00;.41}	.19 (.28) [-.17;.67] {.00;.35}	.18 (.25) [-.48;.58] {.00;.38}	.31 (.34) [-.19;.82] {.00;.48}	.22 (.31) [-.03;.78] {.00;.36}	.21 (.35) [-.32;.77] {.00;.48}	.21 (.32) [-.29;.82] {.00;.40}	.18 (.29) [-.33;.73] {.00;.41}	.25 (.33) [-.31;.72] {.00;.37}
4	.26 (.31) [-.04;.73] {.00;.50}	.19 (.34) [-.48;.80] {.00;.37}	.22 (.29) [-.15;.70] {.00;.45}	.14 (.29) [-.46;.69] {.00;.28}	.25 (.33) [-.30;.80] {.00;.48}	.11 (.26) [-.45;.62] {.00;.25}	.18 (.27) [-.42;.61] {.00;.38}	.16 (.31) [-.54;.79] {.00;.27}	.19 (.29) [-.12;.75] {.00;.28}	.22 (.30) [-.07;.81] {.00;.40}	.17 (.30) [-.55;.67] {.00;.35}	.17 (.29) [-.04;.76] {.00;.32}	.24 (.35) [-.66;.67] {.00;.50}	.19 (.31) [-.48;.74] {.00;.35}	.14 (.22) [-.16;.58] {.00;.27}
5	.24 (.27) [-.25;.66] {.00;.40}	.41 (.38) [-.21;.80] {.03;.63}	.32 (.30) [.00;.75] {.00;.50}	.24 (.32) [-.30;.71] {.00;.50}	.23 (.32) [-.02;.82] {.00;.43}	.18 (.24) [-.21;.62] {.00;.37}	.22 (.26) [-.21;.62] {.00;.44}	.17 (.38) [-.81;.81] {.00;.30}	.19 (.26) [-.13;.73] {.00;.31}	.26 (.32) [.00;.75] {.00;.49}	.20 (.26) [-.01;.66] {.00;.34}	.16 (.30) [-.13;.78] {.00;.27}	.17 (.27) [-.38;.68] {.00;.34}	.16 (.21) [-.23;.62] {.00;.33}	.18 (.29) [-.24;.68] {.00;.42}
6	.43 (.37) [-.27;.81] {.00;.66}	.38 (.25) [-.29;.70] {.26;.52}	.45 (.48) [-.26;.86] {.00;.77}	.34 (.34) [.00;.76] {.00;.56}	.43 (.41) [-.45;.84] {.01;.66}	.50 (.41) [-.46;.82] {.16;.76}	.50 (.35) [-.19;.82] {.30;.65}	.34 (.32) [-.01;.83] {.01;.53}	.39 (.47) [-.21;.88] {.00;.70}	.39 (.38) [.00;.78] {.00;.68}	.43 (.40) [-.27;.83] {.00;.67}	.35 (.40) [-.21;.80] {.00;.62}	.36 (.38) [-.52;.75] {.00;.61}	.62 (.31) [.00;.80] {.51;.76}	.43 (.41) [-.10;.84] {.00;.61}
7	.33 (.36) [-.57;.71] {.00;.57}	.28 (.30) [-.23;.74] {.00;.47}	.27 (.37) [-.44;.80] {.00;.49}	.28 (.34) [-.29;.79] {.00;.49}	.29 (.33) [-.09;.80] {.00;.41}	.22 (.27) [-.12;.67] {.00;.44}	.23 (.30) [-.25;.75] {.00;.46}	.44 (.36) [-.21;.78] {.13;.62}	.19 (.30) [-.35;.75] {.00;.27}	.26 (.30) [-.30;.74] {.00;.46}	.40 (.38) [-.35;.81] {.00;.64}	.30 (.33) [-.01;.80] {.00;.55}	.40 (.35) [-.10;.80] {.12;.61}	.18 (.34) [-.38;.80] {.00;.38}	.20 (.26) [-.15;.72] {.00;.38}
8	.41 (.32) [.00;.79] {.05;.54}	.32 (.30) [.00;.77] {.00;.51}	.30 (.35) [-.27;.78] {.00;.54}	.36 (.35) [.00;.75] {.00;.62}	.55 (.41) [-.15;.85] {.32;.76}	.65 (.37) [.00;.87] {.51;.76}	.27 (.31) [-.25;.79] {.00;.48}	.40 (.35) [.00;.83] {.00;.57}	.34 (.32) [.00;.80] {.00;.54}	.39 (.38) [.00;.78] {.00;.65}	.34 (.39) [-.32;.83] {.00;.58}	.24 (.33) [-.22;.74] {.00;.43}	.31 (.31) [-.09;.80] {.00;.46}	.31 (.38) [-.28;.86] {.00;.45}	.27 (.35) [-.14;.82] {.00;.53}
9	.21 (.28) [-.54;.62] {.00;.42}	.24 (.36) [-.46;.85] {.00;.50}	.19 (.36) [-.46;.75] {.00;.45}	.24 (.42) [-.54;.78] {.00;.50}	.29 (.37) [-.35;.72] {.00;.62}	.15 (.29) [-.28;.83] {.00;.28}	.19 (.29) [-.52;.66] {.00;.41}	.32 (.39) [.00;.79] {.00;.65}	.24 (.28) [.00;.68] {.00;.52}	.29 (.37) [.00;.81] {.00;.45}	.29 (.31) [.00;.69] {.00;.52}	.16 (.25) [-.12;.63] {.00;.28}	.27 (.40) [-.19;.86] {.00;.57}	.28 (.35) [.00;.83] {.00;.53}	.16 (.37) [-.49;.79] {.00;.38}
10	.25 (.41) [-.05;.85] {.00;.40}	.31 (.39) [-.29;.78] {.00;.65}	.24 (.35) [-.10;.76] {.00;.56}	.28 (.43) [-.78;.78] {.00;.57}	.19 (.34) [.00;.84] {.00;.16}	.33 (.40) [.00;.81] {.00;.67}	.29 (.40) [-.30;.90] {.00;.44}	.30 (.44) [-.64;.77] {.02;.62}	.68 (.54) [.00;.91] {.26;.91}	.60 (.58) [-.11;.91] {.00;.89}	.27 (.41) [-.12;.85] {.00;.51}	.60 (.61) [-.22;.91] {.00;.89}	.65 (.56) [-.15;.94] {.16;.82}	.25 (.36) [.00;.76] {.00;.49}	.23 (.40) [-.27;.86] {.00;.44}
11	.44 (.40) [-.06;.83] {.00;.68}	.31 (.35) [-.36;.72] {.00;.56}	.47 (.40) [-.01;.80] {.00;.72}	.59 (.33) [-.15;.84] {.44;.71}	.49 (.40) [-.31;.83] {.34;.57}	.38 (.43) [-.15;.82] {.00;.69}	.37 (.33) [-.46;.75] {.00;.58}	.35 (.39) [-.37;.79] {.00;.65}	.35 (.37) [-.05;.77] {.00;.59}	.35 (.41) [-.42;.83] {.00;.60}	.37 (.40) [-.35;.74] {.00;.69}	.44 (.43) [-.40;.79] {.03;.69}	.39 (.38) [-.30;.74] {.00;.63}	.40 (.35) [-.45;.73] {.00;.63}	.40 (.40) [-.34;.80] {.00;.70}

Bemerkungen: Anordnung: Mittelwert (Standardabweichung) [Minimum; Maximum] {25 %-Quantil; 75 %-Quantil}

Zur Berechnung der Mittelwerte und Standardabweichungen wurden die Trennschärfeparameter der einzelnen Items mittels Fisher-Transformation umgerechnet; die Ergebnisse wurden für die bessere Interpretierbarkeit wieder zurückgerechnet; die Perzentilwerte basieren auf den unveränderten Korrelationen

Der Vergleich der Minima mit den 25 %-Quantilen unter Berücksichtigung der Mittelwerte zeigt, dass meist nur vereinzelt negative Werte auftreten. Für die Subtests 4, 5, 9, 10 und 11 ist dies wegen der vielen unbearbeiteten Items, deren Trennschärfe gleich Null ist, schwieriger einzuschätzen. In Tabelle 10.16 wird daher die Anzahl negativer Trennschärfeparameter aufgeführt. Im Mittel weisen die Subtests 2.5 (0.85) negative Werte auf, wobei in 25 Subtest-Durchführungskonstellationen alle Werte größer Null sind. Für Subtest 10 in Gruppe 8 (Bedingung 5) liegen mit 15 die meisten negativen Werte vor.

Tabelle 10.16: Trennschärfe der Subtests der CV des LPS-neu unterteilt in die 15 Bedingungen der CV

	Einzel									Alle					
	Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar	
	Ohne	Gesamt	Einzel	Ohne	Gesamt	Einzel	Ohne	Gesamt	Einzel	Ohne	Gesamt	Ohne	Gesamt	Ohne	Gesamt
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	.62 (.44) [-10;91] {36;76}	.34 (37) [-53;79] {00;53}	.51 (42) [-33;86] {27;74}	.49 (40) [-10;84] {10;65}	.38 (38) [-38;83] {00;66}	.36 (37) [-15;80] {00;57}	.47 (48) [-07;89] {00;75}	.47 (41) [-23;85] {00;70}	.52 (32) [-10;74] {39;68}	.46 (46) [-45;86] {00;71}	.41 (41) [-38;84] {00;60}	.42 (43) [-40;87] {00;70}	.48 (34) [-14;75] {22;70}	.45 (39) [-15;82] {00;66}	.44 (32) [-07;79] {28;60}
2	.39 (38) [-68;78] {03;62}	.51 (45) [-36;88] {26;73}	.43 (28) [-09;71] {28;62}	.48 (33) [-05;83] {29;63}	.52 (38) [-03;84] {30;78}	.46 (32) [-19;71] {19;65}	.53 (33) [00;82] {33;74}	.51 (37) [-09;85] {26;68}	.53 (28) [-18;79] {38;65}	.35 (36) [-44;71] {00;62}	.53 (29) [00;80] {36;66}	.55 (25) [00;76] {44;62}	.43 (41) [-25;87] {00;62}	.48 (31) [-38;79] {33;62}	.61 (30) [00;85] {45;73}
3	.18 (29) [-33;73] {00;41}	.25 (33) [-31;72] {00;37}	.46 (38) [-39;83] {23;66}	.18 (35) [-34;84] {00;34}	.27 (31) [-07;72] {00;45}	.23 (32) [-16;81] {00;43}	.16 (24) [-11;69] {00;28}	.24 (27) [-09;69] {00;44}	.27 (28) [-07;79] {00;41}	.19 (28) [-17;67] {00;35}	.18 (25) [-48;58] {00;38}	.31 (34) [-19;82] {00;48}	.22 (31) [-03;78] {00;36}	.21 (35) [-32;77] {00;48}	.21 (32) [-29;82] {00;40}
4	.24 (35) [-66;67] {00;50}	.19 (31) [-48;74] {00;35}	.14 (22) [-16;58] {00;27}	.26 (31) [-04;73] {00;50}	.19 (34) [-48;80] {00;37}	.22 (29) [-15;70] {00;45}	.14 (29) [-46;69] {00;28}	.25 (33) [-30;80] {00;48}	.11 (26) [-45;62] {00;25}	.18 (27) [-42;61] {00;38}	.16 (31) [-54;79] {00;27}	.19 (29) [-12;75] {00;28}	.22 (30) [-07;81] {00;40}	.17 (30) [-55;67] {00;35}	.17 (29) [-04;76] {00;32}
5	.16 (30) [-13;78] {00;27}	.17 (27) [-38;68] {00;34}	.16 (21) [-23;62] {00;33}	.18 (29) [-24;68] {00;42}	.24 (27) [-25;66] {00;40}	.41 (38) [-21;80] {03;63}	.32 (30) [00;75] {00;50}	.24 (32) [-30;71] {00;50}	.23 (32) [-02;82] {00;43}	.18 (24) [-21;62] {00;37}	.22 (26) [-81;81] {00;44}	.17 (38) [-13;73] {00;30}	.19 (26) [-01;83] {00;31}	.26 (32) [00;75] {00;49}	.20 (26) [-01;66] {00;34}
6	.43 (40) [-27;83] {00;67}	.35 (40) [-21;80] {00;62}	.36 (38) [-52;75] {00;61}	.62 (31) [00;80] {51;76}	.43 (41) [-10;84] {00;61}	.43 (37) [-27;81] {00;66}	.38 (25) [-29;70] {26;52}	.45 (48) [-26;86] {00;77}	.34 (34) [00;76] {00;56}	.43 (41) [-45;84] {01;66}	.50 (41) [-46;82] {16;76}	.50 (35) [-19;82] {30;65}	.34 (32) [-01;83] {01;53}	.39 (47) [-21;88] {00;70}	.39 (38) [00;78] {00;68}
7	.26 (30) [-30;74] {00;46}	.40 (38) [-35;81] {00;64}	.30 (33) [-01;80] {00;55}	.40 (35) [-10;80] {12;61}	.18 (34) [-38;80] {00;38}	.20 (26) [-15;72] {00;38}	.33 (36) [-57;71] {00;57}	.28 (30) [-23;74] {00;47}	.27 (37) [-44;80] {00;49}	.28 (34) [-29;79] {00;49}	.29 (33) [-09;80] {00;41}	.22 (27) [-12;67] {00;44}	.23 (30) [-25;75] {00;46}	.44 (36) [-21;78] {13;62}	.19 (30) [-35;75] {00;27}
8	.34 (32) [00;80] {00;54}	.39 (38) [00;78] {00;65}	.34 (39) [-32;83] {00;58}	.24 (33) [-22;74] {00;43}	.31 (31) [-09;80] {00;46}	.31 (38) [-28;86] {00;45}	.27 (35) [-14;82] {00;53}	.41 (32) [00;79] {05;54}	.32 (30) [00;77] {00;51}	.30 (35) [-27;78] {00;54}	.36 (35) [00;75] {00;62}	.55 (41) [-15;85] {32;76}	.65 (37) [00;87] {51;76}	.27 (31) [-25;79] {00;48}	.40 (35) [00;83] {00;57}
9	.24 (42) [-54;78] {00;50}	.32 (39) [00;79] {00;65}	-	.29 (31) [00;69] {00;52}	.20 (26) [-12;68] {00;52}	-	.28 (39) [-19;86] {00;57}	.28 (35) [00;83] {00;53}	-	-	-	.19 (33) [-54;79] {00;42}	.26 (37) [-46;85] {00;62}	.17 (32) [-46;83] {00;45}	.19 (29) [-52;66] {00;41}
10	.29 (40) [-30;90] {00;44}	.28 (43) [-78;78] {00;57}	-	.33 (40) [00;81] {00;67}	.29 (42) [-64;85] {00;67}	-	.64 (57) [-22;91] {00;91}	.62 (57) [-15;94] {00;89}	-	-	-	.25 (36) [00;76] {00;49}	.24 (40) [-27;86] {00;44}	.31 (39) [-29;78] {00;65}	.21 (35) [-10;84] {00;56}
11	.31 (35) [-36;72] {00;56}	.47 (40) [-01;80] {00;72}	.59 (33) [-15;84] {44;71}	.49 (40) [-31;83] {34;57}	.38 (43) [-15;82] {00;69}	.37 (33) [-46;75] {00;58}	.35 (39) [-37;79] {00;65}	.35 (37) [-05;77] {00;59}	.35 (41) [-42;83] {00;60}	.37 (40) [-35;74] {00;69}	.44 (43) [-40;79] {03;69}	.39 (38) [-30;74] {00;63}	.40 (35) [-45;73] {00;63}	.40 (40) [-34;80] {00;70}	.26 (31) [-04;73] {00;50}

Bemerkungen: Anordnung: Mittelwert (Standardabweichung) [Minimum; Maximum] {25 %-Quartil; 75 %-Quartil}

Zur Berechnung der Mittelwerte und Standardabweichungen wurden die Trennschärfeparameter der einzelnen Items mittels Fisher-Transformation umgerechnet; die Ergebnisse wurden für die bessere Interpretierbarkeit wieder zurückgerechnet; die Perzentilwerte basieren auf den unveränderten Korrelationen

Andererseits weisen die 75 %-Quartile und die Maxima darauf hin, dass unter Berücksichtigung der Durchführungsbedingungen ein Großteil der Items hinsichtlich der Trennschärfe gute und sehr gute Werte aufweisen. In 58 % der 161 Konstellationen sind 25 % der Werte größer als .50 und immerhin 14 % größer als .70. Der kleinste Wert für eine obere Quartilsgrenze beträgt .16 (Subtest 10, Gruppe 5, Bedingung 15), wobei das Maximum bei .84 und die Quartilsgrenze von Gruppe 3, die ebenfalls Subtest 10 in Bedingung 15 bearbeitet, bei .56 liegen.

Insgesamt zeigen die Trennschärfewerte der CV also, dass es zwar in den einzelnen Bedingungen Items gibt, die nur bedingt funktionieren. In vielen Fällen ergeben sich aber bei Berücksichtigung der Durchführungsbedingung gute Werte.

Tabelle 10.17: Anzahl negativer Trennschärfen der Subtests der CV des LPS-neu unterteilt in die 15 Gruppen der CV

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	M	SD
1	1	4	5	1	3	2	2	2	2	6	3	6	1	3	3	2.93	1.67
2	4	2	1	1	2	0	2	1	5	0	0	1	1	0	2	1.47	1.46
3	1	10	3	2	3	2	1	6	4	2	3	3	4	6	3	3.53	2.33
4	1	8	2	4	3	6	6	5	2	1	2	1	4	2	2	3.27	2.15
5	3	2	0	1	1	3	3	6	2	0	1	3	5	2	1	2.20	1.70
6	1	1	5	0	3	1	1	1	1	0	3	2	2	0	2	1.53	1.36
7	5	4	3	3	2	1	4	1	3	1	2	1	5	7	1	2.87	1.85
8	0	0	2	0	1	0	1	0	0	0	1	3	1	2	3	0.93	1.10
9	5	5	9	7	2	4	4	0	0	0	0	2	1	0	7	3.07	3.03
10	1	1	1	3	0	0	2	15	0	7	1	4	2	0	3	2.67	3.90
11	3	6	1	2	2	3	1	3	1	5	4	8	2	2	3	3.07	1.98

10.4 Vergleich der psychometrischen Indizes von PBV und CV zu den beiden Erhebungszeitpunkten

Hinsichtlich des Vergleichs der mittleren Itemschwierigkeiten der unterschiedlichen Gruppen bzw. Bedingungen stellt sich die Frage, wie groß ein Unterschied zu sein hat, um bedeutsam zu sein. Da hierfür im Kontext der KTT keine Richtlinien und Abhandlungen über die Verteilung von mittleren Itemschwierigkeiten zu finden waren, soll folgende Überlegung die Grenzen von $\pm 5\%$ und $\pm 10\%$ rechtfertigen: Ausgehend von einem beliebigen Wert in der relativen Gesamtgruppe bedeutet eine Veränderung der Schwierigkeit von .05, dass in einer Gruppe von $n = 20$ Versuchspersonen über 40 Items hinweg eine Versuchsperson pro Item besser in der einen bzw. schlechter in der anderen Bedingung abgeschnitten hat (bei kleineren Gruppen oder kleineren Abweichungen entsprechend mehr). Dies wird als noch akzeptable Abweichung im Kontext dieser Bewertung angesehen, insbesondere da die Varianz der Itemschwierigkeiten nicht unerheblich ist, was auf eine gute Verteilung über den Bereich der möglichen Werte schließen lässt.

Der Vergleich der mittleren Itemschwierigkeiten der PBV der beiden Gruppen (vgl. Tabelle 10.12), die die Version entweder zu T1 oder zu T2 bearbeiteten, zeigt, dass sich die Schwierigkeitswerte der Items bei der zweiten Bearbeitung im Mittel nur geringfügig erhöhen (die Items also nur unerheblich leichter werden). Somit wird eine Zusammenfassung der Ergebnisse der PBV unabhängig von der Testposition aus dieser Sicht als gerechtfertigt erachtet. Die Schwierigkeiten der CV zu T1 entsprechen nur für die Subtests 3, 4 und 5 den Werten der PBV zu T1. Der Vergleich der Modi der anderen Subtests hinsichtlich der Testung zu T1 weist konsistent niedrigere Werte in der CV auf, die Items sind in der CV also schwieriger. Für Subtest 10 erweisen sich die Items in der CV im Mittel als konsistent schwieriger. Die Subtests 1 und 2 sind in der CV zu T2 leichter als zu beiden Zeitpunkten in der PBV. Insgesamt entsprechen sich jedoch die mittleren Schwierigkeiten der Subtests in der PBV und der CV zu T2 in noch angemessener Weise; die mittleren Werte der Itemschwierigkeiten der Subtests in der CV zu T1 erweisen sich hingegen mit Ausnahme der Subtests 3, 4 und 5 als durchwegs niedriger.

Die mittleren Itemschwierigkeiten der Subtests 9, 10 und 11 sind in der CV zu T1 durchwegs deutlich geringer als in der PBV zu T1. Für die Subtests 1 und 2 gilt dies ebenso mit Ausnahme

von Gruppe 12 (Subtest 1 in Bedingung 12) und Gruppe 2 (Subtest 2 in Bedingung 3). Die geringsten Werte über alle Subtests hinweg zeigen sich bei Gruppe 6.

In Tabelle 10.13 werden die mittleren Itemschwierigkeiten nach den Durchführungsbedingungen (Tabelle 10.4) sortiert, um deren Effekt veranschaulichen zu können. Bezüglich der Subtests gibt es keine Veränderung. Bedingung 12 (alle Items, nach dem Lösen sichtbar aber nicht mehr änderbar, ohne Zeitanzeige) bedingt im Mittel bei fast allen Subtests in der CV konsistent schwerere Items als in der PBV. Ebenso fällt bei Bedingung 10 (alle Items, nach dem Lösen unsichtbar, ohne Zeitangabe) auf, dass in allen außer den ohnehin relativ unauffälligen Subtests 3, 4 und 5, geringere Itemschwierigkeitswerte auftreten.

Insgesamt deuten die Itemschwierigkeitsindizes darauf hin, dass die Beeinflussung der psychometrischen Eigenschaften der Subtests eher durch den Modus als durch die Erst- oder Zweiddurchführung bedingt ist. Die Subtests 9 und 10, die einen erhöhten Anteil der Speedkomponente aufweisen, sind für einen Modalitätseffekt anfälliger als die übrigen Subtests. Die Subtests 3, 4 und 5, die den höchsten Anteil logischen Denkens und auch Lösungszeit pro Item aufweisen, zeigen sich dagegen am wenigsten durch Modus und Testzeitpunkt hinsichtlich der psychometrischen Kennwerte beeinflusst (vgl. auch Tabelle 10.9 und Tabelle 10.15 hinsichtlich der relativen Konstanz der Trennschärfeindizes der drei Subtests in dem Bereich, in dem die Werte der PBV liegen). Keine Bedingung erzeugte so auffällig abwegige Werte, als dass sie in die folgenden Analysen nicht aufgenommen werden könnte.

10.5 Vergleich der Mittelwerte der Durchführungsbedingungen der CV

In Tabelle 10.18, Tabelle 10.19 und Tabelle 10.20 werden zunächst die Mittelwerte, Standardabweichungen und Gruppengrößen der einzelnen Bedingungskonstellationen sowie deren Zusammenfassungen aufgeführt. Deskriptiv fällt hier bereits auf, dass in manchen Subtests die mittleren Leistungen über die Bedingungen hinweg stark schwanken, während sie sich bei anderen Subtests durchgängig in der Größenordnung der PBV bewegen. Auch die Streuungen schwanken je nach Bedingung für die einzelnen Subtests. Auf den ersten Blick lassen sich keine systematischen Variationen ausmachen. Mit den folgenden inferenzstatistischen Analysen wird dies überprüft.

Tabelle 10.18: Mittelwerte, Standardabweichungen und Fallzahlen je Subtest in jeder Bedingungskonstellation unabhängig vom Erhebungszeitpunkt

Präsen- tation		PBV n=273	Einzel									Alle					
Antwort wort- format			Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigier- bar	
Zeitvor- gabe			Ohne	Ge- samt	Ein- zeln	Ohne	Ge- samt	Ein- zeln	Ohne	Ge- samt	Ein- zeln	Ohne	Ge- samt	Ohne	Ge- samt	Ohne	Ge- samt
Bedin- gung			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Subtest	1	28.33 (9.80)	26.28 (14.6 8, 18)	22.13 (8.18, 15)	27.59 (11.6 2, 17)	21.35 (10.0 5, 17)	25.38 (8.82, 16)	24.00 (8.91, 17)	21.32 (9.08, 19)	24.53 (10.4 5, 15)	20.87 (10.1 3, 15)	25.95 (9.97, 19)	24.53 (10.7 1, 15)	27.93 (13.2 5, 14)	26.63 (11.7 3, 16)	26.35 (9.47, 17)	23.81 (10.1 1, 16)
	2	22.78 (7.57)	20.94 (6.69, 16)	22.11 (8.25, 18)	24.80 (7.73, 15)	23.29 (7.56, 17)	21.18 (8.28, 17)	21.81 (7.15, 16)	19.06 (8.80, 17)	22.16 (8.23, 19)	22.20 (8.39, 15)	18.47 (9.52, 15)	21.79 (8.32, 19)	19.33 (6.67, 15)	24.00 (8.03, 13)	18.24 (10.1 8, 17)	22.94 (9.84, 17)
	3	27.01 (3.73)	26.82 (3.32, 17)	26.38 (3.54, 16)	25.94 (6.11, 18)	27.53 (3.20, 15)	28.00 (3.76, 17)	25.88 (4.26, 17)	27.38 (2.92, 16)	26.29 (3.85, 17)	25.21 (4.18, 19)	25.40 (2.56, 15)	26.43 (2.68, 14)	26.05 (3.56, 19)	26.27 (3.92, 15)	26.85 (3.44, 13)	26.88 (3.28, 17)
	4	16.68 (2.84)	16.94 (3.34, 17)	17.71 (3.70, 17)	15.88 (2.45, 16)	17.17 (3.81, 18)	17.20 (3.08, 15)	17.12 (3.31, 17)	16.82 (1.74, 17)	16.31 (4.09, 16)	15.71 (2.34, 17)	16.95 (2.46, 19)	16.73 (2.79, 15)	17.33 (3.38, 15)	16.79 (2.97, 19)	15.93 (2.38, 15)	17.86 (2.71, 14)
	5	19.06 (3.28)	20.07 (2.50, 14)	19.41 (2.35, 17)	18.94 (2.41, 17)	17.56 (3.20, 16)	19.72 (3.68, 18)	18.13 (5.55, 15)	19.81 (4.34, 16)	19.06 (3.60, 17)	18.38 (3.79, 16)	17.94 (3.44, 16)	18.79 (2.99, 19)	16.60 (2.64, 15)	18.67 (2.67, 15)	18.74 (3.31, 19)	19.60 (3.36, 15)
	6	22.43 (7.70)	16.47 (6.56, 15)	21.07 (6.53, 14)	18.18 (5.54, 17)	18.12 (9.37, 17)	20.38 (7.02, 16)	19.39 (7.10, 18)	21.47 (6.50, 15)	21.75 (7.10, 16)	18.35 (5.74, 17)	20.80 (7.74, 15)	23.94 (7.39, 17)	19.95 (5.59, 19)	18.60 (6.90, 15)	20.93 (5.99, 15)	19.47 (6.54, 19)
	7	30.57 (5.16)	28.79 (4.21, 19)	28.64 (6.12, 14)	30.79 (5.13, 14)	28.53 (4.57, 17)	28.53 (5.81, 17)	30.38 (3.16, 16)	28.72 (4.40, 18)	29.20 (4.46, 15)	30.81 (4.21, 16)	29.08 (5.96, 13)	29.50 (5.04, 14)	29.75 (4.67, 16)	29.79 (4.09, 19)	28.20 (2.74, 15)	31.33 (4.91, 15)
	8	34.43 (5.56)	33.20 (4.87, 15)	33.89 (5.39, 19)	33.57 (5.08, 14)	34.86 (3.51, 14)	33.53 (4.69, 17)	31.12 (6.22, 17)	33.69 (4.24, 16)	32.17 (5.86, 18)	33.00 (5.32, 15)	34.31 (4.99, 16)	32.18 (4.77, 17)	32.14 (8.84, 14)	32.59 (8.35, 17)	32.84 (5.93, 19)	32.80 (6.18, 15)
	9	11.23 (4.33)	15.00 (3.48, 17)	10.67 (5.67, 15)	-	12.08 (4.57, 13)	11.31 (4.05, 29)	-	6.70 (4.31, 30)	7.21 (4.68, 14)	-	-	-	13.63 (3.16, 32)	13.93 (4.26, 30)	14.72 (3.12, 33)	14.72 (3.12, 18)
	10	33.15 (7.22)	29.72 (9.43, 18)	18.38 (4.33, 16)	-	25.59 (6.96, 17)	29.71 (8.48, 28)	-	26.79 (5.68, 28)	26.91 (9.52, 32)	-	-	-	18.79 (5.09, 14)	17.90 (5.47, 31)	18.33 (4.65, 14)	18.33 (4.65, 30)
	11	15.99 (6.96)	13.00 (6.67, 15)	17.38 (8.56, 16)	15.25 (10.3 5, 16)	17.21 (6.99, 14)	13.41 (6.32, 17)	15.28 (6.93, 18)	10.46 (6.57, 13)	15.54 (6.09, 13)	14.13 (7.22, 16)	15.15 (7.48, 13)	17.83 (6.72, 12)	14.29 (7.05, 14)	18.07 (7.02, 14)	14.50 (7.92, 12)	15.82 (7.04, 17)

In Tabelle 10.21 werden die Eintretenswahrscheinlichkeiten der $H_0: \mu_1 = \dots = \mu_m$ für die jeweiligen Haupt- und Interaktionseffekte sowie das klassische und partielle η^2 als Zusammenfassung der elf univariaten Varianzanalysen dargestellt. Die ausführliche Ergebnisdarstellung der Quadratsummen, Freiheitsgrade und F -Werte können in Anhang 15.6 nachgelesen werden. Für den Faktor Bearbeitungszeitpunkt, also ob die CV zu T1 oder zu T2 absolviert wurde, ergeben sich durchgängig statistisch signifikante Effekte; für jeden Subtest ist die mittlere Leistung zu T2 besser. Für alle anderen Haupt- und Interaktionseffekte zeigt sich kein generelles Muster. Zur besseren Bewertung der Unterschiedlichkeiten werden im Folgenden die partiellen η^2 betrachtet.

Tabelle 10.19: Mittelwerte, Standardabweichungen und Fallzahlen je Subtest zusammengefasst für die Bedingungen Antwortformat und Zeitvorgabe

Präsentation		PBV n=273	Einzel			Alle		Einzel			Alle		
Zeitvorgabe/ Antw.format	Zeitvorgabe						Antwortformat						
	Ohne		Gesamt	Einzel	Ohne	Gesamt	Einmal	Weiter	Korr.	Einmal	Weiter	Korr.	
Bedingung			1,4,7	2,5,8	3,6,9	10,12,14	11,13,15	1,2,3	4,5,6	7,8,9	10,11	12,13	14,15
Subtest	1	28.33 (9.80)	22.98 (11.54, 54)	24.04 (9.09, 46)	24.29 (10.44, 49)	26.64 (10.71, 50)	25.00 (10.87, 47)	25.48 (11.96, 50)	23.54 (9.25, 50)	22.16 (9.76, 49)	25.32 (10.71, 34)	27.23 (11.73, 30)	25.12 (10.11, 33)
	2	22.78 (7.57)	21.10 (7.79, 50)	21.83 (8.10, 54)	22.91 (7.70, 46)	18.66 (8.05, 47)	22.78 (8.97, 49)	22.55 (7.62, 49)	22.10 (7.58, 50)	21.14 (8.43, 51)	20.32 (8.32, 34)	21.50 (8.03, 28)	20.59 (9.84, 34)
	3	27.01 (3.73)	27.23 (3.10, 48)	26.90 (3.73, 50)	25.67 (4.85, 54)	26.06 (3.53, 47)	26.54 (3.18, 46)	26.37 (4.49, 51)	27.12 (3.82, 49)	26.23 (3.76, 52)	25.90 (2.68, 29)	26.15 (3.92, 34)	26.87 (3.28, 30)
	4	16.68 (2.84)	16.98 (3.05, 52)	17.08 (3.63, 48)	16.24 (2.76, 50)	16.76 (2.80, 49)	17.08 (2.83, 48)	16.86 (3.25, 50)	17.16 (3.36, 50)	16.28 (2.85, 50)	16.85 (2.79, 34)	17.03 (2.97, 34)	16.86 (2.71, 29)
	5	19.06 (3.28)	19.11 (3.58, 46)	19.40 (3.22, 52)	18.50 (3.98, 48)	17.84 (2.91, 50)	19.00 (3.14, 49)	19.44 (2.40, 48)	18.53 (4.23, 49)	19.08 (3.88, 49)	18.40 (2.99, 35)	17.63 (2.67, 30)	19.12 (3.36, 34)
	6	22.43 (7.70)	18.66 (7.79, 47)	21.07 (6.78, 46)	18.65 (6.09, 52)	20.51 (7.26, 49)	20.71 (6.81, 51)	18.50 (6.33, 46)	19.27 (7.80, 51)	20.46 (6.51, 48)	22.47 (7.39, 32)	19.35 (6.90, 34)	20.12 (6.54, 34)
	7	30.57 (5.16)	28.69 (4.31, 54)	28.78 (5.40, 46)	30.65 (4.11, 46)	29.02 (4.61, 44)	30.19 (4.59, 48)	29.34 (5.09, 47)	29.12 (4.65, 50)	29.55 (4.36, 49)	29.30 (5.04, 27)	29.77 (4.09, 35)	29.77 (4.91, 30)
	8	34.43 (5.56)	33.89 (4.22, 45)	33.20 (5.30, 54)	32.48 (5.58, 46)	33.12 (6.36, 49)	32.51 (6.68, 49)	33.58 (5.04, 48)	33.06 (5.15, 48)	32.92 (5.14, 49)	33.21 (4.77, 33)	32.39 (8.35, 31)	32.82 (6.18, 34)
	9	11.23 (4.33)	10.22 (5.51, 60)	10.16 (4.88, 58)	-	13.98 (3.15, 65)	14.23 (3.86, 48)	12.97 (5.06, 32)	11.55 (4.17, 42)	6.86 (4.38, 44)	-	13.77 (3.70, 62)	14.47 (3.11, 51)
	10	33.15 (7.22)	27.30 (7.32, 63)	26.14 (9.22, 76)	-	17.04 (6.20, 28)	18.11 (5.04, 61)	24.38 (9.35, 34)	28.16 (8.11, 45)	26.85 (7.90, 60)	-	18.18 (5.31, 45)	17.36 (5.57, 44)
	11	15.99 (6.96)	13.62 (7.14, 42)	15.39 (7.17, 46)	14.90 (8.10, 50)	14.64 (6.04, 39)	17.12 (7.42, 43)	15.26 (8.69, 47)	15.18 (6.77, 49)	13.43 (6.85, 42)	16.44 (6.72, 25)	16.18 (7.02, 28)	15.28 (7.04, 29)

Der Messzeitpunkt (CV zu T1 oder T2) stellt für jeden Subtest einen relevanten Effekt dar. Die übrigen Faktoren zeigen für einzelne Subtests Haupteffekte. Die Interaktionen haben nur mehr vereinzelt statistisch bedeutsame Ausmaße. Für die Subtests 6, 8 und 11 zeigen sich überhaupt keine Interaktionen. Die Subtests 2, 4, 7 und 10 weisen je nur eine relevante Interaktion auf. Im Folgenden wird auf den Einfluss der Bedingungsvariationen eingegangen. Die Rolle der Beschaffenheit der Subtests wird erst bei der Interpretation der Äquivalenz weitergehende Beachtung finden.

Tabelle 10.20: Mittelwerte, Standardabweichungen und Fallzahlen je Subtest zusammengefasst für die drei Bedingungen sowie insgesamt

Faktor	Stufen	Bedingung	PBV n=273	PC	Präsentation		Zeitvorgabe			Antwortformat		
					Einzel	Alle	Ohne	Gesamt	Einzel	Einmalig	Weiter	Korr.
					1-9	10-15	1, 4, 7, 10, 12, 14	2, 5, 8, 11, 13, 15	3, 6, 9	1, 2, 3, 10, 11	4, 5, 6, 12, 13	7, 8, 9, 14, 15
Subtest	1		28.33 (9.80)	24.57 (10.58, 246)	23.74 (10.42, 149)	25.85 (10.77, 97)	24.74 (11.25, 104)	24.53 (9.99, 93)	24.29 (10.44, 49)	25.42 (11.41, 84)	24.93 (10.33, 80)	23.35 (9.95, 82)
	2		22.78 (7.57)	21.47 (8.21, 246)	21.92 (7.86, 150)	20.76 (8.73, 96)	19.92 (7.97, 97)	22.28 (8.50, 103)	22.91 (7.70, 46)	21.64 (7.94, 83)	21.88 (7.70, 78)	20.92 (8.97, 85)
	3		27.01 (3.73)	26.47 (3.78, 245)	26.57 (4.03, 152)	26.30 (3.36, 93)	26.65 (3.36, 95)	26.73 (3.47, 96)	25.67 (4.85, 54)	26.20 (3.91, 80)	26.72 (3.87, 83)	26.46 (3.58, 82)
	4		16.68 (2.84)	16.83 (3.02, 247)	16.77 (3.16, 150)	16.92 (2.80, 97)	16.87 (2.92, 101)	17.08 (3.24, 96)	16.24 (2.76, 50)	16.86 (3.05, 84)	17.11 (3.19, 84)	16.49 (2.80, 79)
	5		19.06 (3.28)	18.77 (3.40, 245)	19.01 (3.59, 146)	18.41 (3.06, 99)	18.45 (3.29, 96)	19.21 (3.17, 101)	18.50 (3.98, 48)	19.00 (2.70, 83)	18.19 (3.72, 79)	19.10 (3.65, 83)
	6		22.43 (7.70)	19.91 (6.97, 245)	19.42 (6.94, 145)	20.61 (7.00, 100)	19.60 (7.55, 96)	20.88 (6.76, 97)	18.65 (6.09, 52)	20.13 (7.02, 78)	19.31 (7.41, 85)	20.32 (6.49, 82)
	7		30.57 (5.16)	29.45 (4.64, 238)	29.34 (4.68, 146)	29.63 (4.61, 92)	28.84 (4.43, 98)	29.50 (5.02, 94)	30.65 (4.11, 46)	29.32 (5.04, 74)	29.39 (4.42, 85)	29.63 (4.55, 79)
	8		34.43 (5.56)	33.04 (5.69, 243)	33.19 (5.08, 145)	32.82 (6.50, 98)	33.49 (5.43, 94)	32.87 (5.98, 103)	32.48 (5.58, 46)	33.43 (4.91, 81)	32.80 (6.55, 79)	32.88 (5.56, 83)
	9		11.23 (4.33)	12.10 (4.83, 231)	10.19 (5.18, 118)	14.09 (3.45, 113)	12.18 (4.81, 125)	12.00 (4.87, 106)	-	12.97 (5.06, 32)	12.88 (4.03, 104)	10.95 (5.34, 95)
	10		33.15 (7.22)	23.20 (8.56, 228)	26.67 (8.40, 139)	17.78 (5.42, 89)	24.14 (8.43, 91)	22.57 (8.61, 137)	-	24.38 (9.35, 34)	23.17 (8.46, 90)	22.84 (8.42, 104)
	11		15.99 (6.96)	15.15 (7.28, 220)	14.67 (7.49, 138)	15.94 (6.87, 82)	14.11 (6.61, 81)	16.22 (7.30, 89)	14.90 (8.10, 50)	15.67 (8.04, 72)	15.55 (6.83, 77)	14.18 (6.94, 71)

In Tabelle 10.22 werden die gemittelten η_p^2 der Haupt- und Interaktionseffekte aufgeführt (vgl. Formel (7-3)). Der Bearbeitungszeitpunkt zeigt mit rd. 11 % den größten Effekt, wobei die Intervallgrenzen darauf hinweisen, dass der wahre Wert mit 95-prozentiger Wahrscheinlichkeit größer als .066 und somit sehr bedeutsam ist. Dennoch zeigt sich eine große Streuung über die Subtests hinweg. Die Subtests 1, 2, 3 und 6 haben Werte größer .10, für die Subtests 4, 5 und 10 sind die Werte kleiner .05. Die oberen Konfidenzintervallgrenzen von η_p^2 der Subtests 5 und 10 unterschreiten sogar diese Untergrenze des gemittelten Wertes.

Tabelle 10.21: p -Value, η_p^2 und η_c^2 der Haupt- und Interaktionseffekte der univariaten Varianzanalysen der elf Subtests der LPS-neu hinsichtlich der Faktoren Präsentation, Antwortformat, Zeitvorgabe und Bearbeitungszeitpunkt

		1	2	3	4	5	6	7	8	9	10	11
Präsentation	p ,	.04,	.64,	.30,	.89,	.02,	.66,	.64,	.14,	<.01,	<.01,	.04,
	η_p^2 ,	.018,	.001,	.004,	<.001,	.022,	.001,	.001,	.009,	.217,	.291,	.019,
	η_c^2	.012	<.001	.003	<.001	.018	<.001	<.001	.008	.163	.251	.015
Antwortformat	p ,	.16,	.72,	.73,	.28,	.03,	.82,	.97,	.58,	<.01,	.02,	.50,
	η_p^2 ,	.015,	.003,	.003,	.011,	.028,	.002,	<.001,	.005,	.104,	.036,	.007,
	η_c^2	.010	.002	.002	.009	.022	.001	<.001	.004	.068	.023	.005
Zeitvorgabe	p ,	.62,	.02,	.08,	.13,	.03,	.22,	.01,	.22,	.08,	.02,	.01,
	η_p^2 ,	.004,	.031,	.021,	.017,	.031,	.013,	.037,	.013,	.013,	.025,	.039,
	η_c^2	.003	.022	.017	.014	.024	.010	.032	.011	.008	.016	.033
Zeitpunkt	p ,	<.01,	<.01,	<.01,	<.01,	.01,	<.01,	<.01,	<.01,	.08,	.03,	<.01,
	η_p^2 ,	.253,	.254,	.127,	.047,	.025,	.128,	.083,	.095,	.053,	.013,	.094,
	η_c^2	.215	.233	.113	.040	.020	.117	.074	.089	.033	.008	.083
Präsentation × Antwortformat	p ,	.33,	.42,	.25,	.22,	.42,	.35,	.88,	.40,	<.01,	.60,	.86,
	η_p^2 ,	.009,	.007,	.012,	.012,	.007,	.009,	.001,	.008,	.086,	.001,	.001,
	η_c^2	.006	.005	.009	.010	.006	.007	<.001	.007	.055	<.001	.001
Präsentation × Zeitvorgabe	p ,	.43,	.08,	.19,	.87,	.12,	.24,	.13,	.68,	.63,	.69,	.74,
	η_p^2 ,	.003,	.012,	.007,	<.001,	.010,	.006,	.010,	.001,	.001,	.001,	.001,
	η_c^2	.002	.009	.006	<.001	.008	.005	.008	<.001	<.001	<.001	<.001
Präsentation × Zeitpunkt	p ,	.31,	.34,	.83,	.21,	.95,	.58,	.02,	.64,	<.01,	.44,	.26,
	η_p^2 ,	.004,	.004,	<.001,	.006,	<.001,	.001,	.025,	.001,	.062,	.003,	.006,
	η_c^2	.003	.003	<.001	.005	<.001	.001	.021	<.001	.039	.002	.005
Antwortformat × Zeitvorgabe	p ,	.17,	.64,	.68,	.63,	.05,	.13,	.90,	.67,	<.01,	<.01,	.60,
	η_p^2 ,	.026,	.010,	.010,	.011,	.038,	.029,	.005,	.010,	.039,	.095,	.013,
	η_c^2	.017	.007	.008	.009	.031	.024	.004	.009	.024	.064	.010
Antwortformat × Zeitpunkt	p ,	.35,	.95,	.08,	.24,	.17,	.92,	.22,	.94,	.12,	.22,	.59,
	η_p^2 ,	.009,	<.001,	.021,	.012,	.015,	.001,	.013,	.001,	.018,	.013,	.005,
	η_c^2	.006	<.001	.017	.010	.012	<.001	.011	<.001	.011	.008	.004
Zeitvorgabe × Zeitpunkt	p ,	<.01,	.34,	.75,	<.01,	.52,	.50,	.94,	.94,	.52,	.24,	.17,
	η_p^2 ,	.049,	.009,	.002,	.040,	.005,	.006,	.001,	.001,	.002,	.006,	.017,
	η_c^2	.032	.006	.002	.034	.004	.005	<.001	<.001	.001	.004	.014
Präsentation × Antwortformat × Zeitvorgabe	p ,	.07,	.20,	.64,	.16,	.53,	.87,	.51,	.72,	.93,	.19,	.21,
	η_p^2 ,	.022,	.014,	.004,	.015,	.005,	.001,	.006,	.003,	<.001,	.008,	.014,
	η_c^2	.015	.009	.003	.013	.004	<.001	.005	.002	<.001	.005	.012
Präsentation × Antwortformat × Zeitpunkt	p ,	.16,	.57,	.04,	.47,	.06,	.35,	.23,	.84,	.29,	.15,	.64,
	η_p^2 ,	.015,	.005,	.027,	.006,	.023,	.009,	.013,	.001,	.005,	.009,	.004,
	η_c^2	.010	.003	.022	.005	.018	.007	.011	.001	.003	.006	.003
Präsentation × Zeitvorgabe × Zeitpunkt	p ,	.64,	.44,	.19,	.85,	.94,	.90,	.64,	.47,	.95,	.77,	.23,
	η_p^2 ,	.001,	.002,	.007,	<.001,	<.001,	<.001,	.001,	.002,	<.001,	<.001,	.007,
	η_c^2	<.001	.002	.006	<.001	<.001	<.001	<.001	.002	<.001	<.001	.006
Antwortformat × Zeitvorgabe × Zeitpunkt	p ,	.17,	.59,	.56,	.47,	<.01,	.26,	.62,	.85,	.32,	.53,	.88,
	η_p^2 ,	.027,	.012,	.012,	.015,	.072,	.022,	.011,	.006,	.010,	.006,	.006,
	η_c^2	.017	.008	.010	.012	.059	.018	.010	.005	.006	.003	.005
Präs. × Antw.f. × Zeitvorgabe × Zeitpunkt	p ,	.03,	.42,	.89,	.13,	.39,	.68,	.95,	.26,	.23,	.88,	.51,
	η_p^2 ,	.029,	.007,	.001,	.017,	.008,	.003,	<.001,	.012,	.006,	<.001,	.006,
	η_c^2	.019	.005	<.001	.014	.006	.003	<.001	.010	.004	<.001	.005
Gesamt	η_c^2	.366	.315	.218	.176	.231	.200	.179	.150	.414	.389	.201

Bemerkung: Effekte mit $p < .10$ sind grau hinterlegt.

Der zweite sehr hohe Wert findet sich für den Haupteffekt Präsentation ($\eta_p^2(\text{Präs.}) = .053$); hierbei ist allerdings die noch höhere Varianz, die sich in den beiden extrem hohen η_p^2 in den Subtests 9 und 10 (vgl. Tabelle 10.21) begründet, zu beachten. Zudem hängt es vom Subtest ab, welche Bedingung (einzeln vs. alle) die höheren Leistungen provoziert: In den Subtests 1, 9 und 11 erzielen die Probanden, die alle Items sehen, bessere Leistungen; in den Subtests 5 und 10 die

Probanden, die die Items einzeln sehen (vgl. Tabelle 10.20). Diese Umstände begründen wohl auch, dass die Untergrenze des Konfidenzintervalls nur marginal größer Null ist.

Für den Haupteffekt Zeitvorgabe zeigen sich zwar niedrige Werte, die aber konsistent größer .01 sind; im Mittel klärt der Faktor 2.2 % Varianz, sieben der elf Subtest weisen signifikante Mittelwertsunterschiede auf. Nur auf Subtest 1 scheinen die Variationen der Zeitvorgabe keinen Einfluss zu haben. Aus Tabelle 10.20 ist ersichtlich, dass es keine Systematik dahingehend gibt, dass in einer Bedingung für jeden Subtest bessere oder schlechtere Werte erbracht werden. Zwei Trends sind jedoch auszumachen. Die Anzeige der Gesamtzeit führt, außer in den Subtests 9 und 10, nicht zu den niedrigsten, sondern in fünf Fällen zu den höchsten und in vier Fällen zu den zweithöchsten Leistungen. Die Begrenzung der Bearbeitungszeit auf 10 Sekunden führt in fünf der neun angewandten Subtests zu den niedrigsten Rohwerten, in den Subtests 2 und 7 hingegen zu den höchsten.

Der Faktor Antwortformat hat im Mittel eine noch geringere Effektgröße von .019, die Varianz ist jedoch deutlich größer als bei der Zeitvorgabe. Bei den meisten Subtests spielt das Antwortformat kaum eine Rolle. Lediglich die Subtests 5, 9 und 10 weisen bedeutsam von Null verschiedene Effekte auf. Mit 10.4 % Varianzklärung ist die Relevanz für Subtest 9 am höchsten, wobei – ebenso wie bei Subtest 10 – die Möglichkeit, ein Item nur einmal beantworten zu können, zu den höchsten Ergebnissen führt. Die Korrekturoption hingegen produziert die geringsten Werte. Bei Subtest 5 führt die „weiter“-Option zu einer deutlichen Verringerung der mittleren Leistungen (vgl. Tabelle 10.20).

Tabelle 10.22: Mittelwert, Standardabweichung, und Intervallgrenzen des 90 %-Konfidenzintervalls der partiellen η^2 bezüglich der elf Subtests des LPS-neu

	$M(\eta_{p(\text{Effekt})}^2)$	$SD(\eta_{p(\text{Effekt})}^2)$	$u(\eta_{p(\text{Effekt})}^2)$	$o(\eta_{p(\text{Effekt})}^2)$
Präsentation	.053	.101	.003	.103
Antwortformat	.019	.030	.004	.035
Zeitvorgabe	.022	.011	.017	.028
Zeitpunkt	.107	.082	.066	.147
Präsentation × Antwortformat	.014	.024	.002	.026
Präsentation × Zeitvorgabe	.005	.004	.003	.007
Präsentation × Zeitpunkt	.010	.019	.001	.019
Antwortformat × Zeitvorgabe	.026	.026	.013	.039
Antwortformat × Zeitpunkt	.010	.007	.006	.013
Zeitvorgabe × Zeitpunkt	.013	.017	.004	.021
Präsentation × Antwortformat × Zeitvorgabe	.008	.007	.005	.012
Präsentation × Antwortformat × Zeitpunkt	.011	.008	.007	.015
Präsentation × Zeitvorgabe × Zeitpunkt	.002	.003	.000	.003
Antwortformat × Zeitvorgabe × Zeitpunkt	.018	.019	.009	.028
Präs. × Antw.f. × Zeitvorgabe × Zeitpunkt	.008	.009	.004	.012

Die einzige Interaktion, deren Konfidenzintervall über dem Wert für einen relevanten Effekt liegt, betrifft die Faktoren Antwortformat und Zeitvorgabe (90 %-KI von $M(\eta_{p(\text{Antw.f.} \times \text{Zeitv.})}^2) =$

[.013; .039], vgl. Abbildung 10.2). Dies liegt insbesondere an Subtest 10, aber auch an den Subtests 5 und 9, weshalb der Haupteffekt Antwortformat auch unter diesem Gesichtspunkt relativiert werden muss. Für Subtest 5 begründet sich die Interaktion alleine in einer geringeren Leistung der Bedingungskonstellation Weiter-Button-Klicken und keine Zeitangabe. Für die Subtests 9 und 10 zeigt sich, dass die Probanden, die keine Zeitvorgaben sahen und jedes Item einzeln nur einmalig bearbeiten konnten, im Schnitt die höchsten Werte erzielten; demgegenüber erreichten Probanden im selben Antwortformat aber mit Einblendung der Gesamtzeit die geringste Leistung.

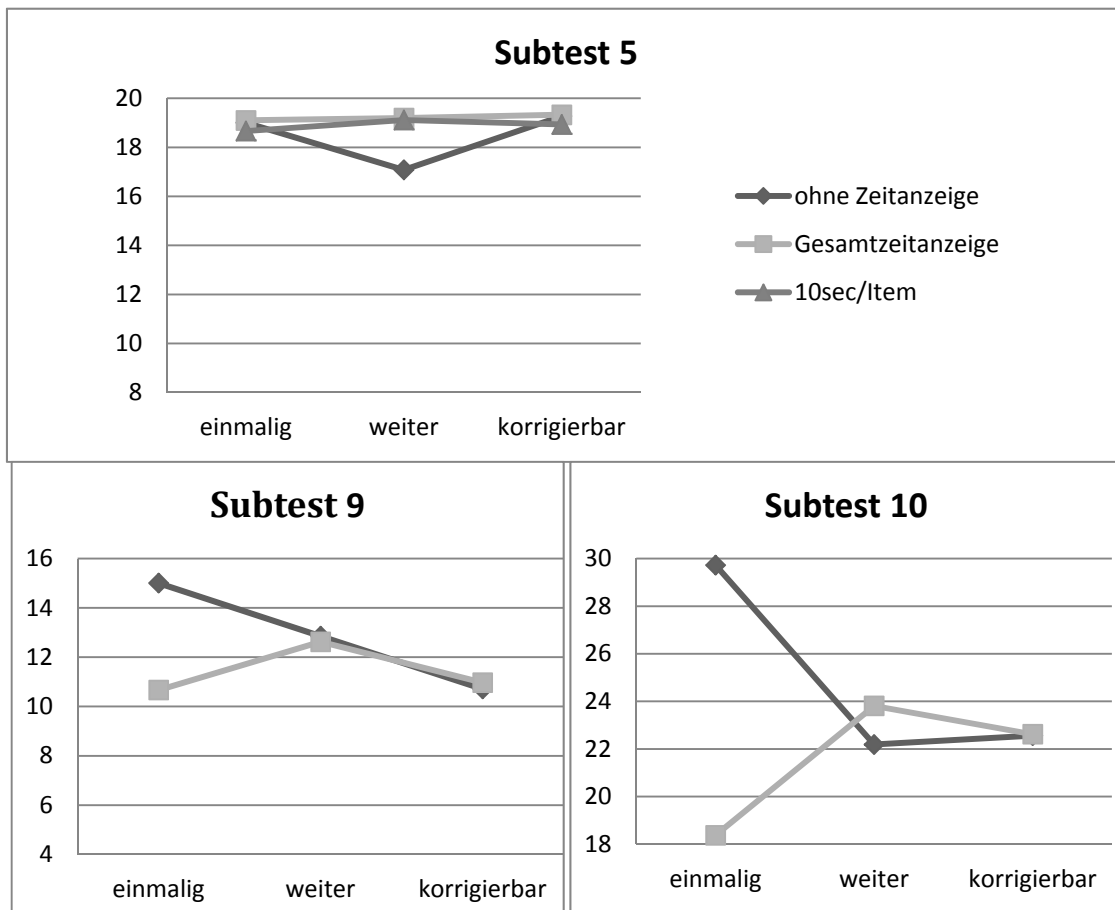


Abbildung 10.2: Darstellung der statistisch signifikanten Interaktion Antwortformat × Zeitvorgabe der Subtests 5, 9 und 10

Bei der gemeinsamen Betrachtung von Zeitvorgabe und Bearbeitungszeitpunkt liegen zwei, allerdings ebenfalls konträre Interaktionen für die Subtests 1 und 4 vor (vgl. Abbildung 10.3). Bei Subtest 1 wird der starke Haupteffekt des Erhebungszeitpunktes der CV zu T1 oder T2 deutlich. Zudem zeigen sich bei der Testung zu T2 die besten Werte, wenn keine Zeitangaben gemacht werden. Findet die Durchführung hingegen zu T1 statt, erzielen Probanden in dieser Bedingung im Mittel die niedrigsten Ergebnisse. Während bei Subtest 4 bei keiner oder der Anzeige der Gesamtzeit nur der Haupteffekt des Erhebungszeitpunktes greift, führt eine Begrenzung der Bearbeitungszeit auf 10 Sekunden zu einer Egalisierung dieses Unterschiedes. Über alle Subtests

hinweg ist die Höhe der Effektes jedoch mit relativer Sicherheit marginal größer als Null (90 %-KI von $M(\eta^2_{p(\text{Zeitv.} \times \text{Zeitp.})}) = [.004; .021]$).

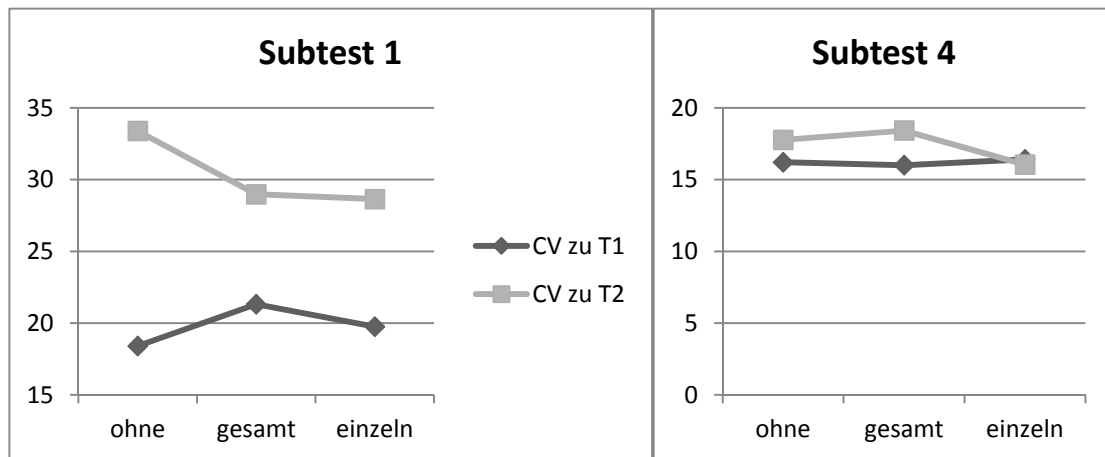


Abbildung 10.3: Darstellung der statistisch signifikanten Interaktion Zeitvorgabe × Zeitpunkt der Subtests 1 und 4

Die beiden signifikanten Interaktionen der Subtests 7 und 9 hinsichtlich der Darstellungsart der Items und des Bearbeitungszeitpunktes begründen sich in beiden Fällen auf einem geringeren Unterschied der Leistungen zu T1 und T2, wenn die Items alle zusammen präsentiert werden. Hierbei übertrifft die mittlere Anzahl richtig gelöster Items zu T1 bei Subtest 9 sogar die Zahl der Lösungen zu T2, wenn die Items einzeln dargestellt werden. Insofern ist der Haupteffekt des Erhebungszeitpunktes für Subtest 9 zu relativieren, da er nur im schlechteren Abschneiden der Probanden, die die Items zu T1 einzeln bearbeitet haben, begründet ist.

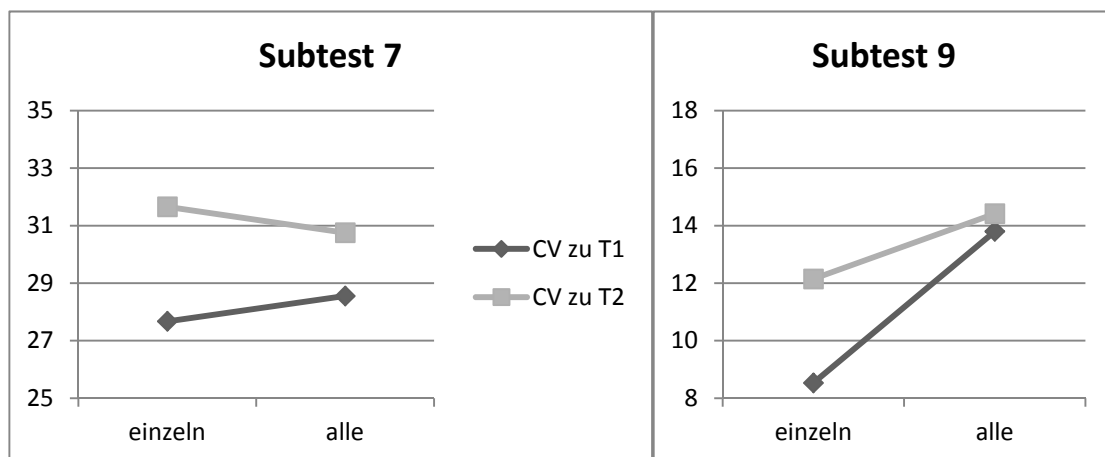


Abbildung 10.4: Darstellung der statistisch signifikanten Interaktion Präsentation × Zeitpunkt der Subtests 7 und 9

Die beiden Interaktionen Präsentation × Zeitvorgabe von Subtests 2 und Präsentation × Antwortformat von Subtest 9 (vgl. Abbildung 10.5) stellen für die Bedingungskonstellationen jeweils eine Ausnahme dar, da für die restlichen Subtests so gut wie keine Effekte beobachtbar sind. Bei Subtest 2 führt die Gesamtpräsentation der Items ohne visuelle Zeitvorgabe zu ei-

nem verringerten Mittelwert der Leistungen. Für Subtest 9 stellt die korrigierbare Version für die ohnehin zu geringeren Leistungen führende Einzelpräsentation ein weiteres Hemmnis dar. Über alle Subtests hinweg zeigen sich jedoch keine weiteren derartigen Effekte von relevantem Ausmaß.

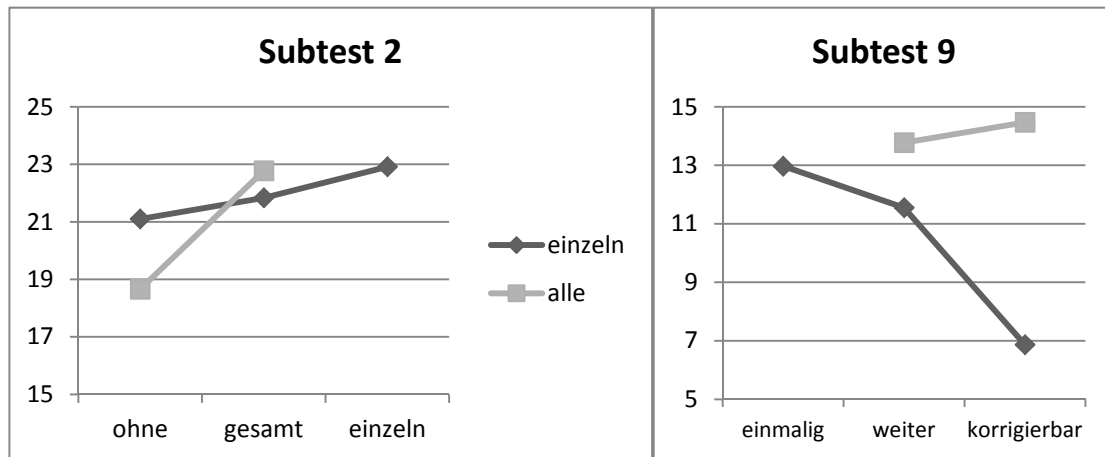


Abbildung 10.5: Darstellung der statistisch signifikanten Interaktion Präsentation × Zeitvorgabe von Subtests 2 und Präsentation × Antwortformat von Subtest 9

Die letzte statistisch signifikante 2-fach Interaktion Antwortformat × Zeitpunkt von Subtests 3 weist entsprechend dem geringen Effekt von 2.1 % (80 %-KI von $\eta^2_{p(\text{Antw.f.} \times \text{Zeitp.})} = [0; .045]$ nur geringe Abweichungen bzgl. des bereits festgestellten Haupteffekts des Erhebungszeitpunktes auf (vgl. Abbildung 10.6). Die Bedingung, dass nach Anklicken der Lösung das nächste Item folgt, führt nur bei der Bearbeitung des Subtests 3 zu T2 zu einer geringeren mittleren Leistung. Zu T1 ergeben sich lediglich marginale Unterschiede.

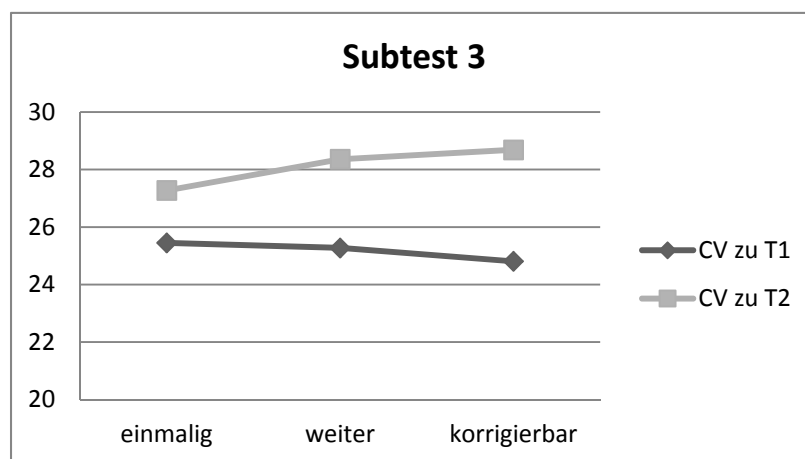


Abbildung 10.6: Darstellung der statistisch signifikanten Interaktion Antwortformat × Zeitpunkt von Subtest 3

Insgesamt bleibt also festzuhalten, dass über die Subtests hinweg nur die Haupteffekte statistische Relevanz aufweisen. Die vereinzelt signifikanten Interaktionen können im Kontext der meist elffachen Testung entweder als zufällig angesehen oder einer dem jeweiligen Subtest in-

härenten Eigenschaft zugeschrieben werden. Auf die Besonderheiten der Subtests wird im Folgenden im Kontext der Äquivalenz noch eingegangen, so dass dieser Abschnitt mit der Feststellung beendet werden muss, dass die Realisierung möglichst unterschiedlicher Bedingungen der CV nur als bedingt erfolgreich gewertet werden kann.

10.6 Vergleich der Mittelwerte der PBV und der CV

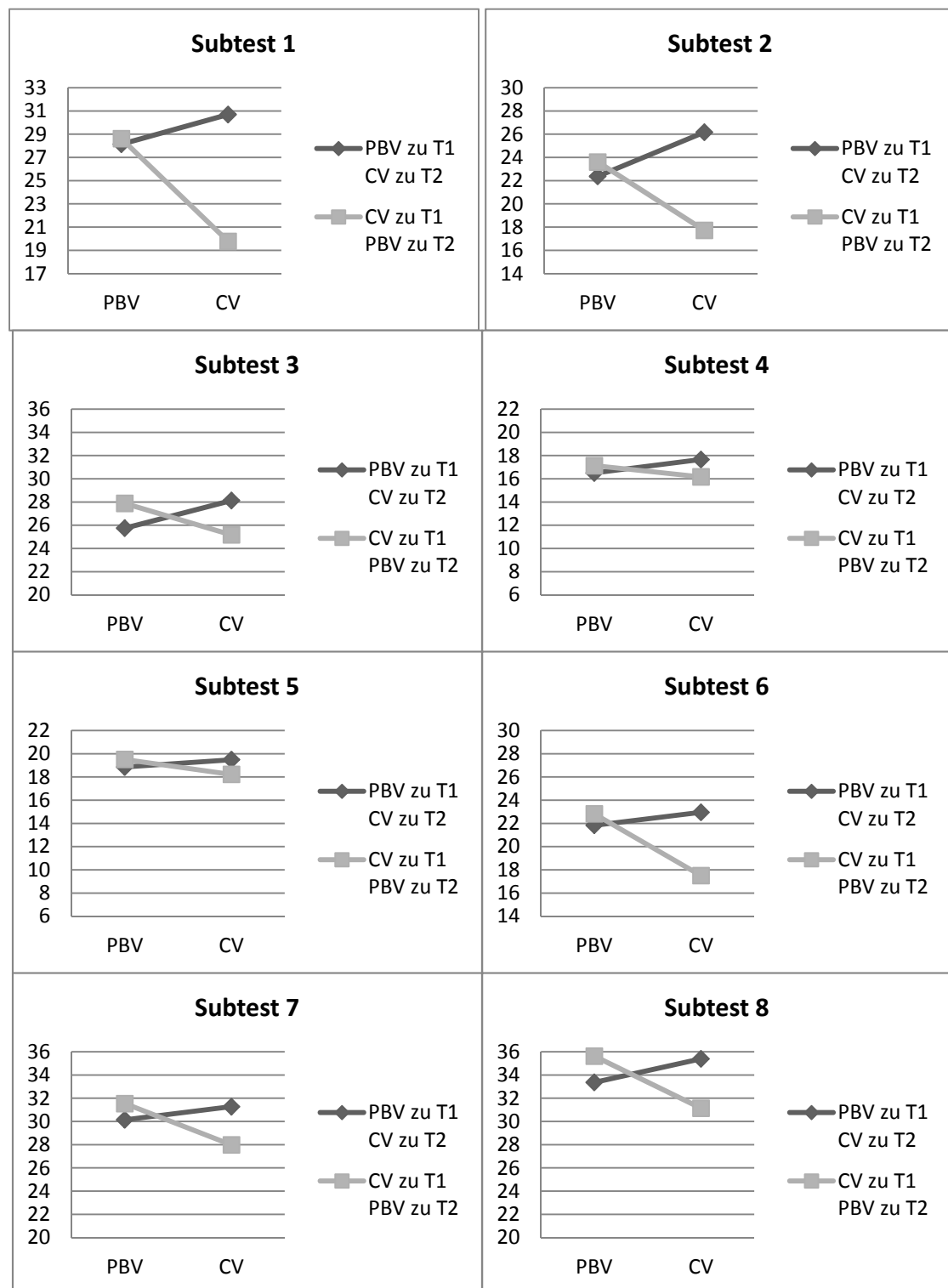
Tabelle 10.23 zeigt deutlich, dass für alle Subtests eine Interaktion bezüglich der Reihenfolge (zuerst CV oder zuerst PBV) und dem Modus der Durchführung (PBV oder CV) zusätzlich zum Haupteffekt des Modus vorliegt. Lediglich für die Subtests 3 und 4 liegt kein Haupteffekt der Durchführungsbedingung vor. Die ausführlichen Darstellungen der Mittelwerte, Streuungen und Teilstichprobengrößen werden in Anhang 15.7, die Kennwerte der Varianzanalysen in Anhang 15.8 bereitgestellt.

Tabelle 10.23: Effektstärken der Haupt- und Interaktionseffekte der univariaten Varianzanalysen der elf Subtests der LPS-neu hinsichtlich der Faktoren Modus, Präsentation, Antwortformat, Zeitvorgabe und Bearbeitungszeitpunkt

Effekt	Subtest											M	SD
	1	2	3	4	5	6	7	8	9	10	11		
Modus	.162	.043	.004	.002	.023	.139	.073	.091	.067	.676	.048	.121	.191
Modus × Bearbeitungszeitpunkt	.378	.494	.369	.151	.102	.259	.238	.333	.064	.074	.137	.236	.143
Modus × Präsentation	<.001	.003	.006	.006	.001	<.001	.004	.013	.204	.316	.003	.051	.107
Modus × Antwort	.019	.010	.001	.010	.016	.001	.007	.018	.119	.040	.012	.023	.034
Modus × Zeit	.015	.012	.024	.011	.048	.005	.005	.027	.012	.018	.015	.017	.012
Modus × Zeitpunkt × Präsentation	<.001	.023	.001	<.001	<.001	<.001	.007	<.001	.021	.004	<.001	.005	.009
Modus × Zeitpunkt × Antwort	<.001	.020	.006	.014	.011	.014	.003	.004	.002	.008	.003	.008	.006
Modus × Zeitpunkt × Zeit	.029	.009	.030	.029	.006	.009	.015	.031	.002	.003	.005	.015	.012
Modus × Präsentation × Antwort	.003	.002	.029	<.001	.011	.029	.009	.001	.026	.001	<.001	.010	.012
Modus × Präsentation × Zeit	.002	<.001	<.001	<.001	.001	.009	<.001	<.001	.004	.012	.003	.003	.004
Modus × Antwort × Zeit	.007	.011	.033	.011	.018	.009	.008	.004	.021	.120	.017	.023	.033
Modus × Zeitpunkt × Präs. × Antwort	.003	.007	.008	.023	.006	.002	<.001	.008	.010	<.001	.014	.007	.007
Modus × Zeitpunkt × Präs. × Zeit	<.001	.010	<.001	<.001	<.001	<.001	<.001	.004	<.001	.018	<.001	.003	.006
Modus × Zeitpunkt × Antwort × Zeit	.009	.012	.036	.024	.074	.007	.015	.045	.003	.002	.019	.022	.022
Modus × Präs. × Antwort × Zeit	.009	.011	.004	.002	.003	.008	.002	<.001	<.001	<.001	.004	.004	.004
Modus × Zeitpkt. × Präs. × Antw. × Zeit	.003	.017	.010	<.001	.007	.022	.006	.005	.005	<.001	<.001	.007	.007

Bemerkung: Hinterlegte Felder entsprechen Signifikanzniveau von $\alpha < .15$

Zunächst werden die Interaktionen zwischen Modus und Bearbeitungsreihenfolge betrachtet, auch wenn sie prinzipiell nicht gänzlich unbeeinflusst von 3- und 4-fach Interaktionen interpretiert werden können. Dennoch stellt diese Interaktion den einzigen Effekt dar, der für jeden Subtest Relevanz hat. Inwiefern die Richtung konsistent ist, wird durch Abbildung 10.7 veranschaulicht.



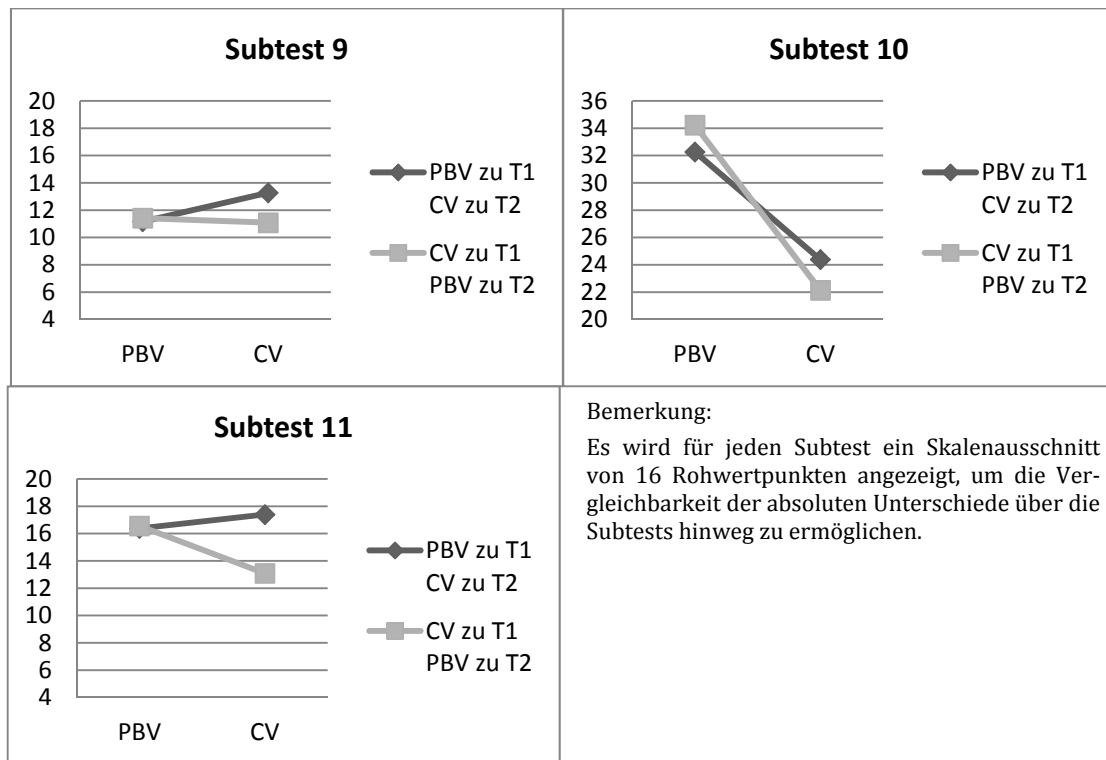


Abbildung 10.7: Interaktion zwischen dem Modus (PBV und CV) und der Bearbeitungsreihenfolge der elf Subtests des LPS-neu

Generell ähneln sich die Ergebnisse der Berufsschulstichprobe und der studentischen Stichprobe (vgl. Abbildung 9.3). Die zweite Testung führt mit einer Ausnahme (Subtest 10) zu einem besseren Ergebnis. Für die Subtests 1, 2, 6 und 11 ergeben sich für die zweite Testung als PBV im Mittel dieselben Ergebnisse wie für die PBV zu T1, während die CV zu T2 zu einer geringeren Verbesserung des ersten Resultats führt als der Vergleich der CV zu T1 mit der PBV zu T2. Die zweite Testung als PBV ergibt für die Subtests 3, 4, 5, 7 und 8 bessere Werte als die erste Testung als PBV. Ebenso liegt eine Steigerung der Leistung der zweiten Testung als CV für diese Subtests vor. Die beiden Ergebnisse der Gruppe, die zuerst die CV und dann die PBV von Subtest 9 bearbeiteten, unterscheiden sich im Mittel nicht, lediglich die CV zu T2 zeigt einen erhöhten Wert. Subtest 10 ist der einzige Subtest, bei dem die CV unabhängig vom Bearbeitungszeitpunkt niedrigere Werte liefert als die PBV. Die zweite Testung führt aber dennoch jeweils zu einem höheren Ergebnis.

Mit Ausnahme von Subtest 9 führt die Testung als PBV im Mittel zu besseren Ergebnissen als die Testung in der CV (vgl. auch Tabelle 10.20). Mit Nachsicht bezüglich dieses „Alles über einen Kamm scheren“ werden zuletzt die Ergebnisse einer multivariate Varianzanalyse nur mit den Faktoren Modus und Bearbeitungszeitpunkt vorgestellt. Als Testparameter wurde den Ausführungen Olsons (1976) folgend mit dem Pillai Trace V die robusteste der von SPSS berechneten Teststatistiken ausgewählt. Multivariat bestätigt sich die Interaktion der beiden Faktoren ($V = 0.76$, $F(11, 193) = 56.71$, $p < .01$, $\eta_p^2 = .76$). Auch die Haupteffekte bezüglich der Bearbeitungsreihenfolge ($V = 0.15$, $F(11, 193) = 3.16$, $p < .01$, $\eta_p^2 = .15$) und Modus ($V = 0.63$, $F(11, 193) = 30.20$, $p < .01$, $\eta_p^2 = .63$) werden multivariat statistisch signifikant. Dabei ist zu beachten, dass die Voraussetzung der Gleichheit der Kovarianzmatrizen verletzt ist ($Box - M =$

434.05, $F(253,123586) = 1.52$, $p < .01$), wobei dies höchsten hinsichtlich der Effektgröße der Reihenfolge Relevanz hat, da die beiden anderen Effekte insgesamt zu deutlich vorliegen.

Die zweite Testung hat – erwartungsgemäß – somit immer einen positiven Effekt hinsichtlich der Höhe der mittleren Rohwertsumme. Je nach Subtest und Modus fällt dieser Effekt jedoch unterschiedlich aus. Für die Subtests 1, 2, 6, 7, 8, 9 und 11 scheint die zweite Testung als PBV einen deutlich stärkeren Anstieg in der Leistung zu verursachen als die zweite Testung als CV. Es kann daraus geschlossen werden, dass die CV im Vergleich zur PBV in diesen Tests einen negativen Einfluss auf die Leistung der Probanden hat. Für Subtest 10 wird dieser negative Einfluss eindeutig. Dennoch zeigt sich auch hier der Vorteil der zweiten Testung. Nur die Subtests 3, 4 und 5 weisen lediglich sehr schwache Tendenzen in der oben beschriebenen Form auf. Für diese Subtests ist der Moduseffekt im Vergleich zu den übrigen Subtests gering (vgl. Tabelle 10.23). Mit Ausnahme der Subtests 7 und 11 zeigen sich für die Subtests weitere bedeutsame Interaktionen mit der Durchführungsbedingung. Diese werden nun vor der endgültigen Diskussion der Rolle der Subtesteigenschaften berichtet.

Weitere 2-fach-Interaktionen mit dem Durchführungsmodus liegen insbesondere für die Subtest 8, 9 und 10 vor (Tabelle 10.23). Abbildung 10.8 macht den Unterschied des Einflusses auf die drei Subtests deutlich. Während der geringe Effekt von $\eta_p^2 = .013$ bei Subtest 8 in der grafischen Aufbereitung kaum ersichtlich ist, zeigt sich, dass die Präsentation aller Items in der CV für Subtest 9 förderlich ist, für Subtest 10 hingegen ein noch größeres Hindernis darstellt als die Bearbeitung der Items einzeln in der CV.

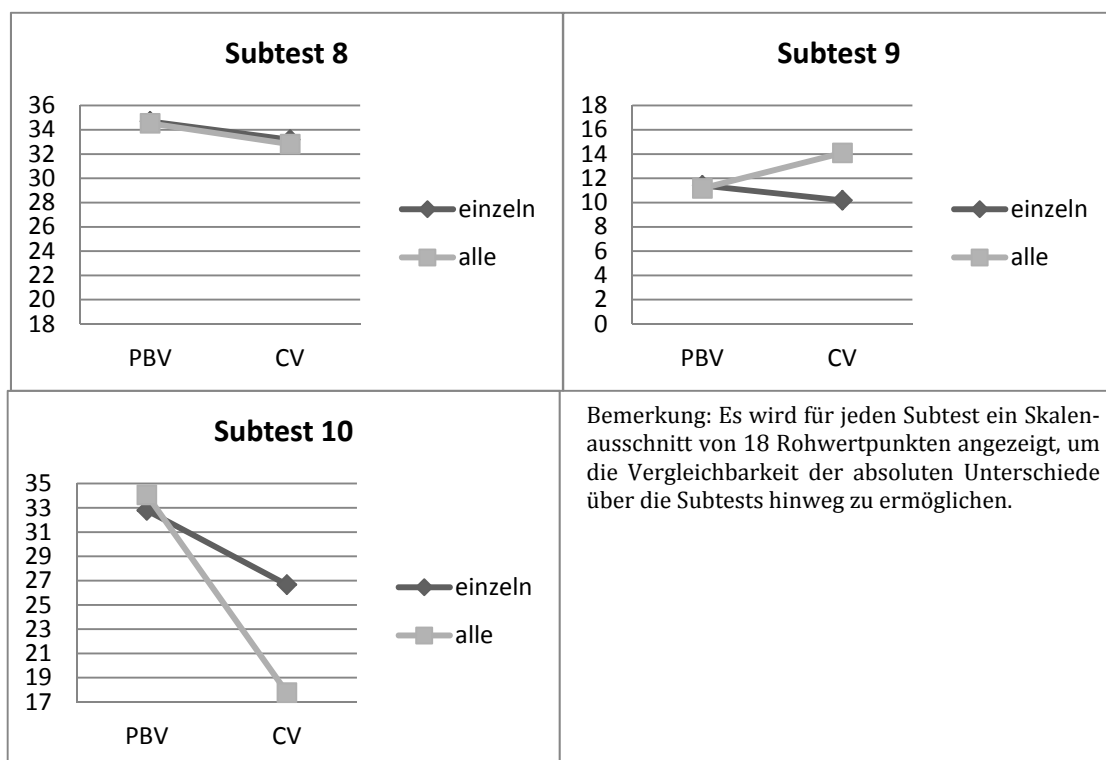


Abbildung 10.8: Interaktion zwischen dem Modus (PBV und CV) und der Präsentationsart der Subtests 8, 9 und 10 des LPS-neu

In Abbildung 10.9 wird die Interaktion von Modus und Antwortformat der CV dargestellt. Für Subtest 1 entsprechen sich die Gruppen in der PBV in etwa, in der CV resultieren jedoch für die Versuchspersonen, die Korrekturmöglichkeiten hatten, die niedrigsten Werte, während sich die Versuchspersonen in den beiden anderen Bedingungen etwas weniger verschlechterten. Bei Subtest 8 zeigen die Probanden, die in der CV keine Änderungen ihrer Eingaben mehr machen durften, in beiden Modi dieselbe Leistung. Die anderen sind in der PBV etwas besser, haben dafür in den CV etwas geringere Werte. In Subtest 9 zeigen sich konstante Ergebnisse für die Korrekturbedingung über die Modi. In den beiden anderen Antwortformaten weisen die Versuchspersonen höhere Werte in der CV auf. Für Subtest 10 zeigt sich eine generelle Verschlechterung durch die Bearbeitung am Computer; die Interaktion wird durch höhere Abnahme der Versuchspersonen in der „weiter“-Bedingung begründet.

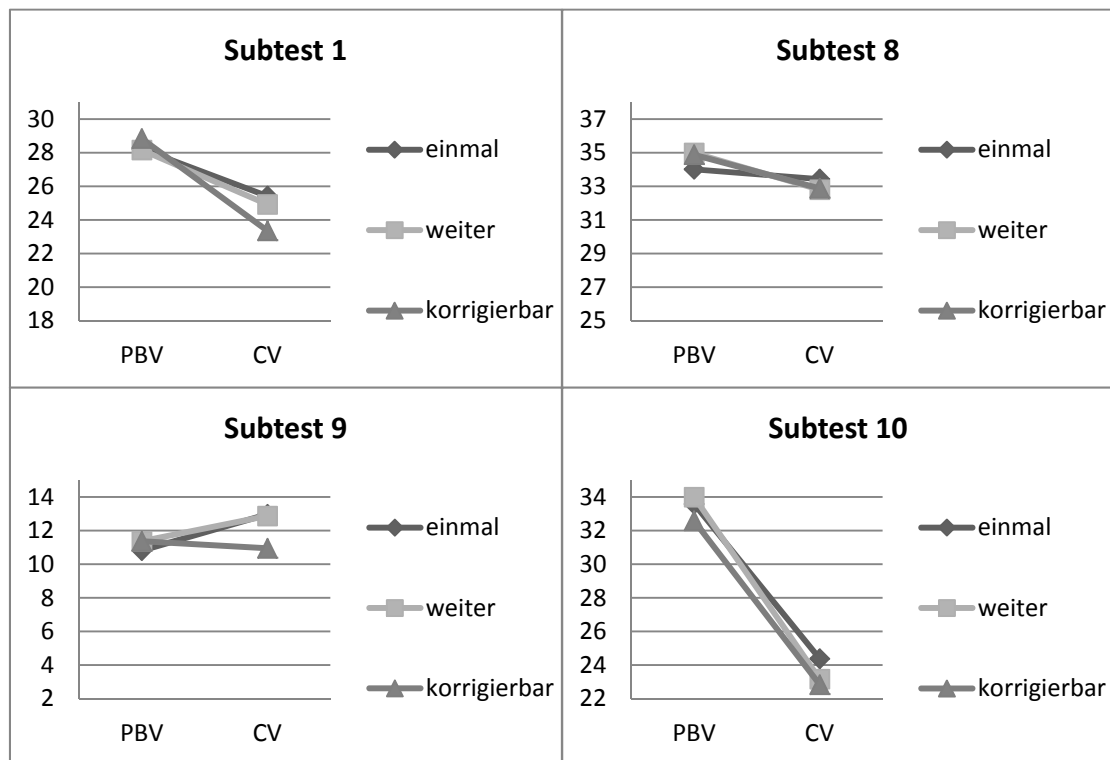


Abbildung 10.9: Interaktion zwischen dem Modus (PBV und CV) und dem Antwortformat der Subtests 1, 8, 9 und 10 des LPS-neu (Skalierung auf 12 Punkte)

Mit Ausnahme von Subtest 10 zeigt sich bei den signifikanten Interaktionen bezüglich Modus und Zeitvorgabe, dass keine Zeitangabe oder die Anzeige der Gesamtzeit zu annähernd gleichen Ergebnissen in den beiden Modi führt (vgl. Abbildung 10.10). Die Vorgabe von 10 Sekunden pro Item führt in den Subtests 3, 5 und 8 zu relativen Verschlechterungen der CV im Kontext der beiden anderen Bedingungen. Die Interaktion in den Subtests 9 und 10 liegt in der Überkreuzung der beiden Reihenfolgen innerhalb der Modi begründet. Die Anzeige der Gesamtzeit führt jeweils zu etwas schlechteren Ergebnissen, wobei diese Effekte relativ gering sind (vgl. Tabelle 10.23).

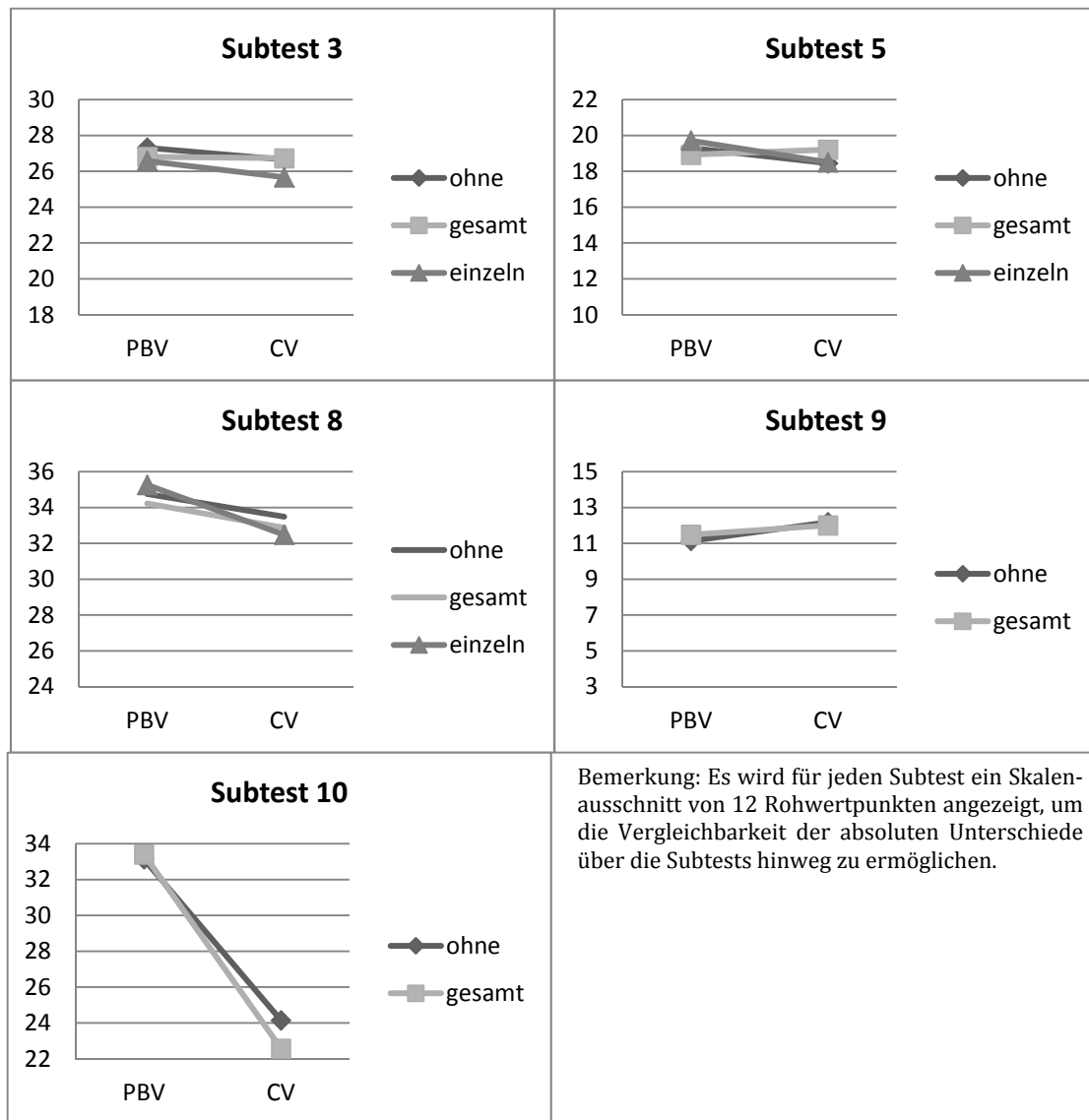


Abbildung 10.10: Interaktion zwischen dem Modus (PBV und CV) und der Zeitvorgabe der Subtests 3, 5, 8, 9 und 10 des LPS-neu

Die 3-fach-Interaktion zwischen dem Modus, der Bearbeitungsreihenfolge und der Präsentationsart bei Subtest 2 unterscheidet sich grafisch nur geringfügig von der 2-fach Interaktion Modus \times Bearbeitungsreihenfolge (vgl. Abbildung 10.7 und Abbildung 10.11). In der CV zu T1 ist der Unterschied zur PBV bei der Einzelitempräsentation geringer als bei der Gesamtpräsentation. Für die CV zu T2 zeigen sich fast keine Unterschiede. Die PBV von Subtest 9 unterscheidet sich in den vier Gruppen fast nicht. Werden in der CV alle Items gezeigt, ist es egal, ob der Subtest zum T1 oder T2 bearbeitet wurde. Die Einzelpräsentation zu T1 führt zu einem schlechteren Ergebnis als die PBV. Zwischen PBV und der Einzelpräsentation zu T2 besteht hingegen fast kein Unterschied.

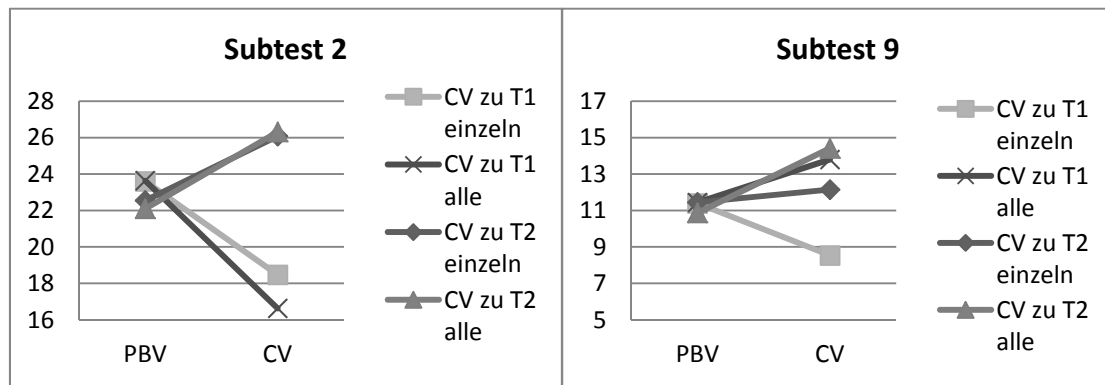


Abbildung 10.11: Interaktion zwischen dem Modus (PBV und CV), der Bearbeitungsreihenfolge und der Präsentationsart der Subtests 2 und 9 des LPS-neu (Skalierung auf 12 Punkte)

Die einzige signifikante 3-fach-Interaktion zwischen dem Modus, der Bearbeitungsreihenfolge und dem Antwortformat liegt für Subtest 2 vor. Diese Interaktion ist jedoch vorwiegend in unterschiedlichen Leistungen der Probanden in der PBV zu T2 begründet, während sie sich in der CV zu T1 nur geringfügig unterscheiden. Wurde die PBV zu T1 durchgeführt, zeigt sich lediglich der Bearbeitungseffekt hinsichtlich einer Verbesserung bei der zweiten Testung.

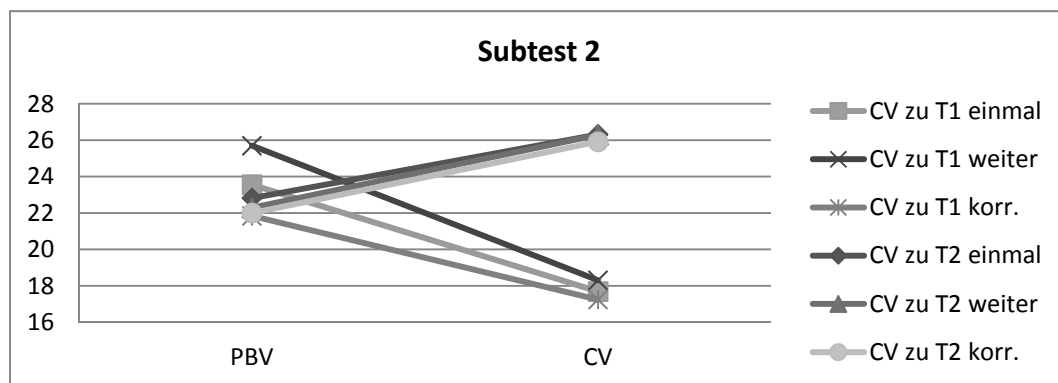


Abbildung 10.12: Interaktion zwischen dem Modus (PBV und CV), der Bearbeitungsreihenfolge und dem Antwortformat von Subtest 2 des LPS-neu (Skalierung auf 12 Punkte)

Die Interaktionen zwischen Modus, Bearbeitungsreihenfolge und Zeitvorgabe der Subtests 1, 3, 4 und 8 deuten alle die 2-fach-Interaktion Modus \times Reihenfolge an (Abbildung 10.13). Für Subtest 1 zeigt sich für die PBV eine stichprobenbedingte auffällige Streuung. Die Steigerung der Leistung ist in der Gruppe der Versuchspersonen, die die CV zu T1 bearbeitete, relativ identisch; wurde die CV zu T1 ohne Zeitangabe bearbeitet, ist sie etwas größer als bei Bearbeitung mit Anzeige der Gesamtzeit. Die Steigerung der Leistung in der Gruppe, die die CV zu T2 bearbeiteten, ist etwas geringer, insbesondere wenn die CV mit einer Zeitbegrenzung von 10 Sekunden pro Item absolviert wurde. Die Subtests 3 und 4 zeigen ähnliche Verläufe, wenngleich die Unterschiede der sechs Gruppen nicht so gravierend ausfallen. In Subtest 3 sind alle Gruppen, die die PBV zu T2 absolvierten, im Mittel besser als die Gruppen der PBV zu T1. Die Interaktion begründet sich vorwiegend durch das relativ schlechtere Abschneiden der Versuchspersonen, die die

CV zu T2 mit Zeitbegrenzung pro Item bearbeiteten; zudem „überholten“ die Versuchspersonen, die die Gesamtzeit angezeigt bekamen, die Versuchspersonen ohne Zeitangabe. Bei den Subtests 4 und 8 fällt je nur eine Gruppe aus dem in der 2-fach-Interaktion festgestellten Zusammenhang. Versuchspersonen, die die CV zu T2 mit Einzelzeitbegrenzung bearbeiteten, blieben in ihrer Leistung lediglich stabil.

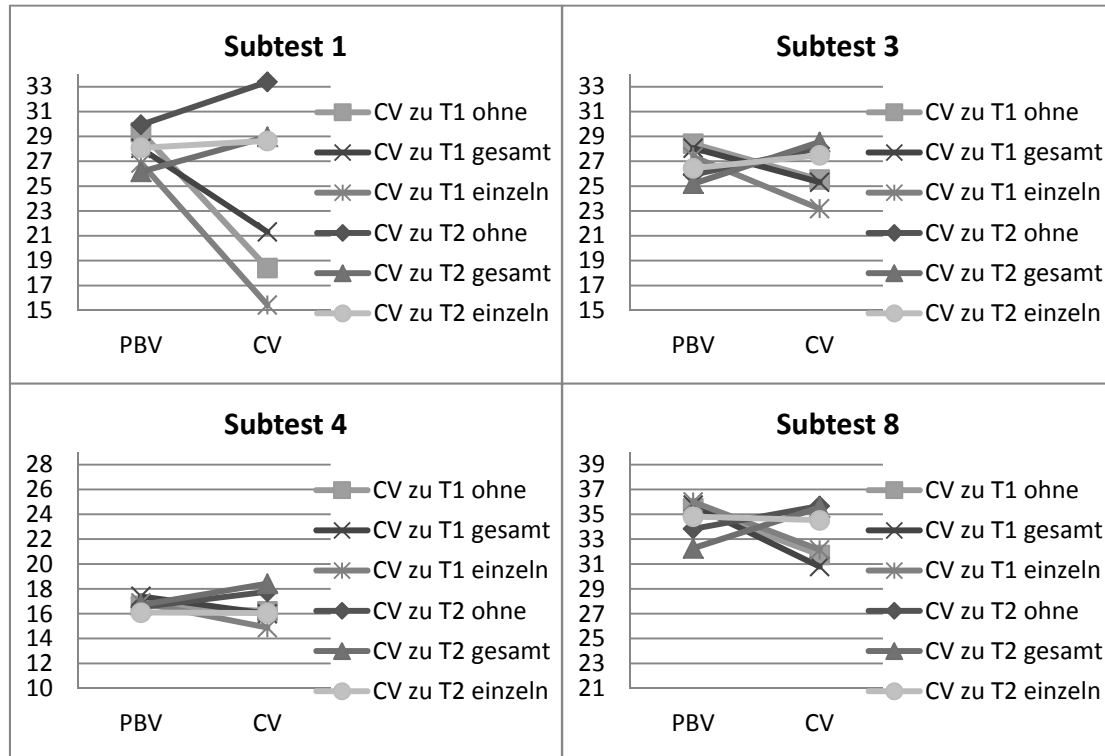


Abbildung 10.13: Interaktion zwischen dem Modus (PBV und CV), der Bearbeitungsreihenfolge und der Zeitvorgabe der Subtests 1, 3, 4 und 8 des LPS-neu (Skalierung auf 19 Punkte)

Für Subtest 3 zeigen sich im Vergleich zu den Subtests 6 und 9 nur geringe Schwankungen, die die signifikante Interaktion zwischen den Faktoren Modus, Präsentation und Antwortformat begründen können (Abbildung 10.14). Werden die Items von Subtest 3 einzeln und nur einmalig bearbeitbar präsentiert, ergeben sich für die CV leicht höhere Werte als für die PBV, für die fünf anderen Gruppen ist dies umgekehrt. Die leicht unterschiedlichen Differenzen von PBV und CV begründen wohl den Effekt. Für Subtest 6 weisen alle sechs Gruppen bessere Werte in der PBV auf; werden die Items in der CV jedoch alle auf einmal präsentiert und bleiben sie korrigierbar, ist die Differenz relativ geringer. Insgesamt zeigt sich eine starke stichprobenbedingte Streuung in der PBV. Bei Subtest 9 fällt die mittlere Leistung der CV, wenn die Items einzeln präsentiert werden, allerdings korrigierbar bleiben, im Vergleich zur PBV ab. Werden die Items einzeln präsentiert und die Versuchspersonen genötigt, einen „weiter“-Button zu klicken, bleibt sie in etwa gleich. Für die restlichen drei Gruppen ist die Leistung in der CV hingegen besser als in der PBV.

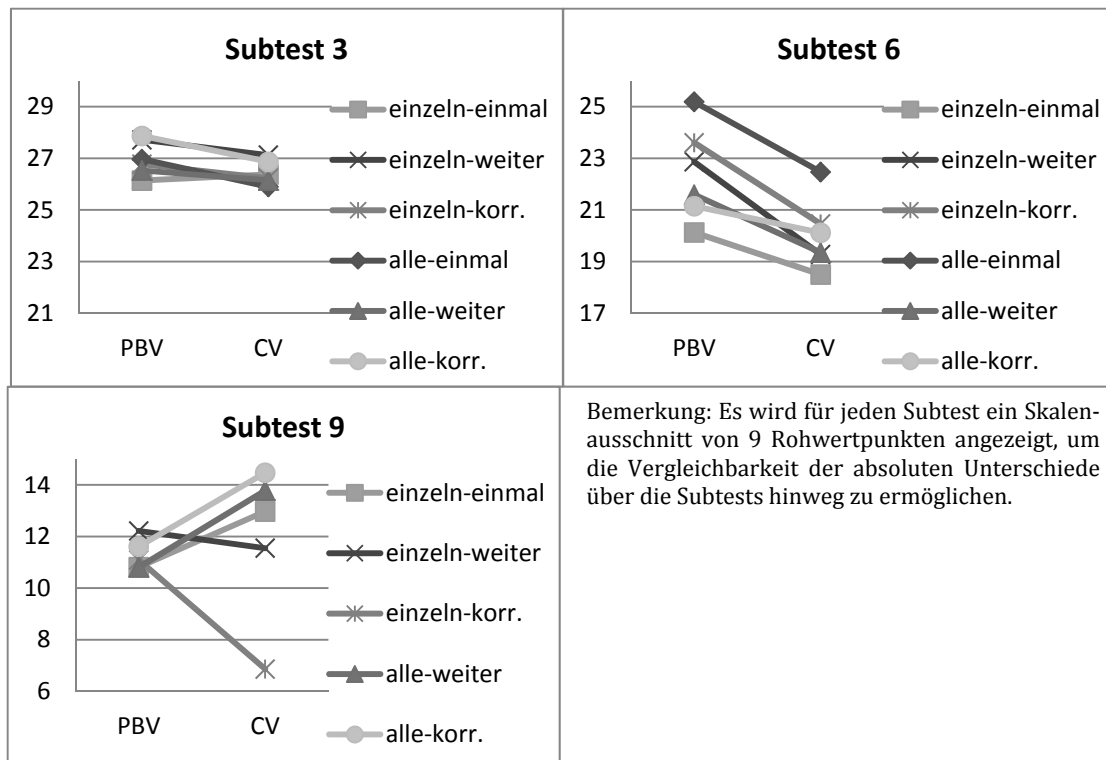


Abbildung 10.14: Interaktion zwischen dem Modus (PBV und CV), der Präsentation und dem Antwortformat der Subtests 3, 6 und 9 des LPS-neu

Die Interaktion Modus \times Präsentation \times Zeitvorgabe (Abbildung 10.15) des Subtests 10 begründet sich in dem umgekehrten Verhältnis der Leistungen innerhalb der Ausprägungen einzeln und alle. Werden die Items einzeln präsentiert, zeigt sich ein kleiner Vorsprung, wenn keine Zeit angezeigt wird. Bei der Präsentation aller Items auf einmal ist der mittlere Rohwert etwas höher, wenn die Gesamtzeit angezeigt wird. Allerdings begründet sich dieser Effekt auf den Unterschied in der PBV; zudem ist er mit der kleinste signifikante Effekt ($\eta_p^2 = .012$).

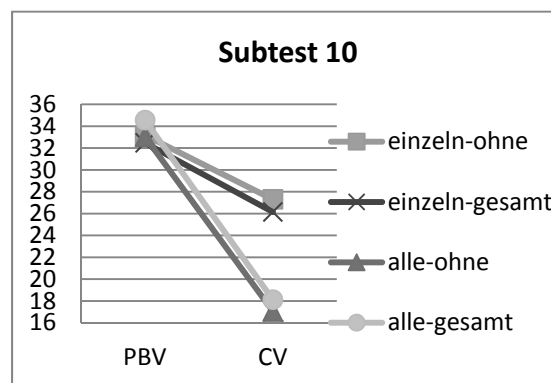


Abbildung 10.15: Interaktion zwischen dem Modus (PBV und CV), der Präsentation und der Zeitvorgabe von Subtest 10 des LPS-neu

Die wohl komplizierteste signifikante 3-fach-Interaktion liegt zwischen Modus, Antwortformat und Zeitvorgabe bei den Subtests 3, 9 und 10 vor. Allerdings zeigt Abbildung 10.16 deutlich, dass

die Effekte durch ein jeweils unterschiedliches Zusammenwirken der Faktoren bedingt sind und sich nur vereinzelt Analogien zwischen zwei der drei Subtests finden lassen. Daher soll die ausführliche Beschreibung dieser, ebenso wie der höheren Interaktionen außen vor bleiben. Hervorgehoben wird lediglich eine gemeinsame Besonderheit von Subtest 9 und 10: Die einmalige Bearbeitung eines Items ohne Anzeige der Zeit führt jeweils zum besten Ergebnis in der CV; das schlechteste Ergebnis der CV wird hingegen erbracht, wenn die Gesamtzeit angezeigt wird. Spannend hierbei ist zudem, dass die Versuchspersonen, die in Subtest 9 diesen besten Mittelwert erzeugen (einmal – ohne), in Subtest 10 den niedrigsten Mittelwert aufweisen (einmal – gesamt, vgl. Tabelle 10.4). Ein Stichprobeneffekt wird somit höchst unwahrscheinlich. Im Folgenden wird aufgrund der Höhe des Effekts exemplarisch die 4-fach-Interaktion Modus \times Reihenfolge \times Antwortformat \times Zeitvorgabe in Abbildung 10.17 dargestellt und erläutert.

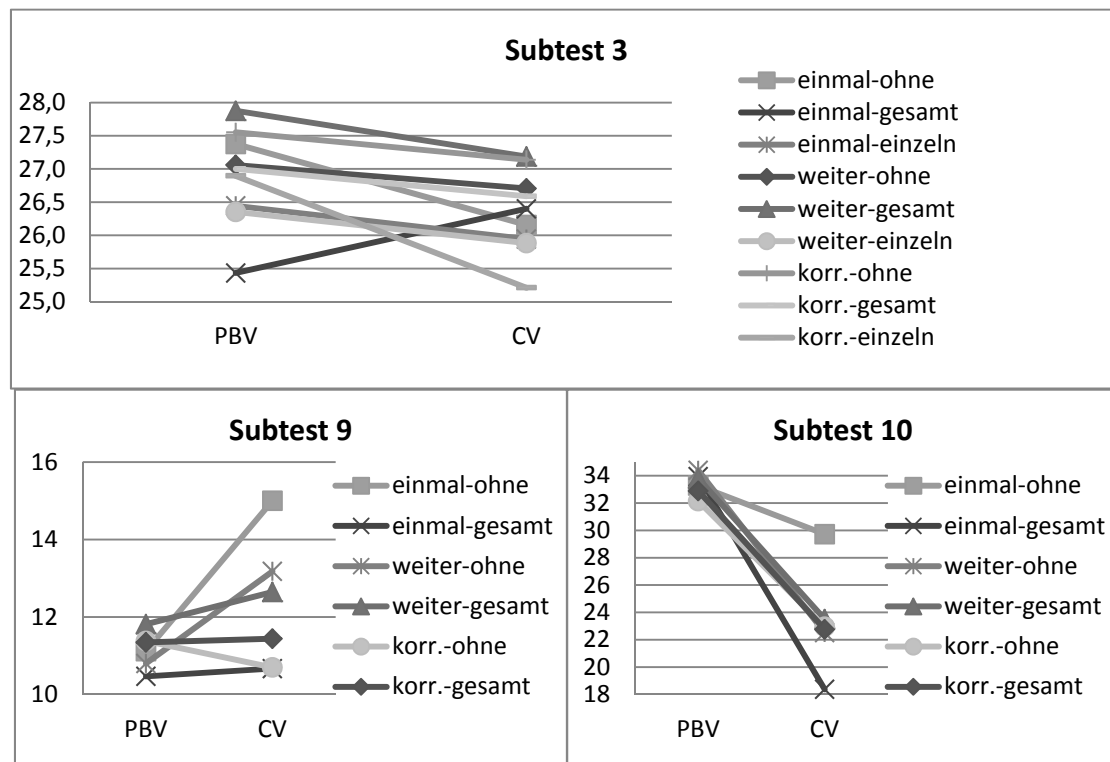


Abbildung 10.16: Interaktion zwischen dem Modus (PBV und CV), dem Antwortformat und der Zeitvorgabe der Subtests 3, 9 und 10 des LPS-neu

Eine 4-fach-Interaktion ist nie leicht zu interpretieren, insbesondere wenn nicht die Überprüfung einer konkreten Hypothese im Vordergrund steht. In den Gruppen ohne Anzeige der Zeit scheint für Subtest 3 in der Interaktion zwischen Modus und Bearbeitungsreihenfolge vorwiegend die zweite Testung zu einer Leistungssteigerung zu führen. Ist die Bearbeitungszeit hingegen auf 10 Sekunden pro Item begrenzt, fallen die Leistungen in der CV geringer aus und kommen über das Niveau der PBV zu T1 nicht hinaus. In den Gruppen, die die Gesamtzeit angezeigt bekommen, unterscheiden sich die Leistungen in der PBV insgesamt sehr auffällig. Lässt man diesen Stichprobenunterschied außer Acht, zeigt sich dieselbe Struktur, wie in den Gruppen ohne Zeitanzeige. Ähnliches gilt für Subtest 8 in den Bedingungen ohne und mit Gesamtzeitanzeige. In Subtest 5 fällt in den Bedingungen ohne Zeitanzeige lediglich die Gruppe „CV zu T2 - weiter“

aus dem Rahmen, die in beiden Modi unterdurchschnittliche Werte erzielte. Die Ergebnisse dieser Gruppe („CV zu T2 – weiter“) sind mit dieser und der Einzelitemzeitbegrenzung über alle anderen Subtests und Bedingungen konstant: Die zweite Testung als CV führt zu etwas besseren Werten als die erste Testung. Dasselbe Muster findet sich für die Gruppe „CV zu T2 – korrigierbar“; die Ausnahme stellt hier Subtest 8 mit Zeitbegrenzung pro Item dar. Die Antwortbedingung „einmal“ bei der CV zu T2 zeigt bei Begrenzung der Bearbeitungszeit pro Item für alle Subtests keinen Anstieg der Leistung, während dies bei den anderen Bedingungen in unterschiedlichem Ausmaß beobachtbar ist. Wurde die CV zu T1 in der Antwortbedingung „weiter“ bearbeitet, zeigt sich für jeden Subtest und jede Zeitbedingung ein Anstieg in der zweiten Testung als PBV. Dies gilt mit Ausnahme von Subtest 5 bei der Zeitangabe gesamt ebenso für die beiden anderen Antwortformate. Die Ausnahme zeigt sich in gleichen Mittelwerten bei der CV und der PBV.

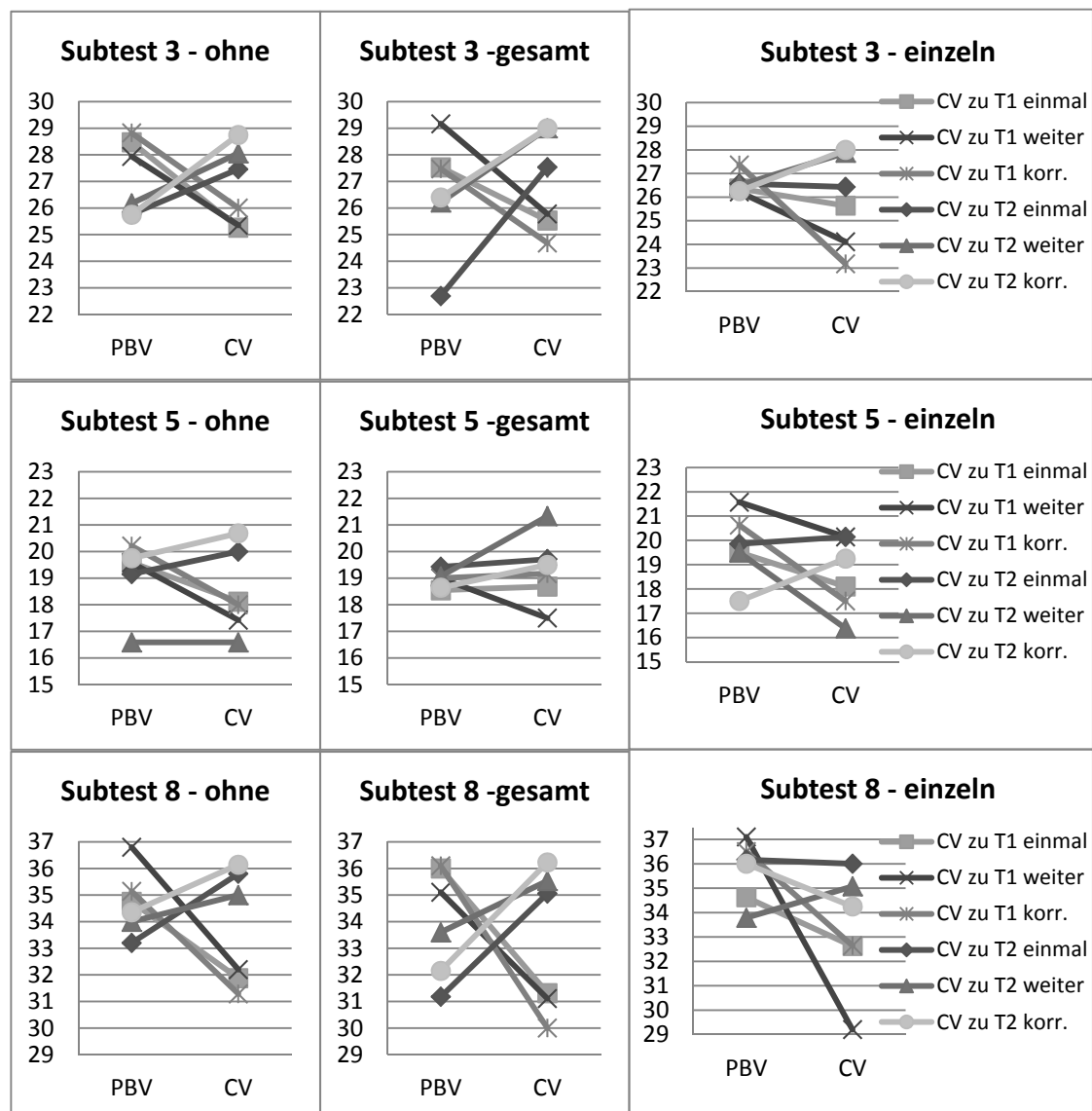


Abbildung 10.17: Interaktion zwischen dem Modus (PBV und CV), der Bearbeitungsreihenfolge, dem Antwortformat und der Zeitvorgabe der Subtests 3, 5 und 8 des LPS-neu (Skalenausschnitt von 8 Rohwertpunkten)

Insgesamt zeigt diese Interaktion, dass vor allem das Fehlen der Anzeige der Gesamtzeit zu Ergebnissen führt, die der 2-fach-Interaktion Modus \times Bearbeitungszeitpunkt am nächsten kommen. Die Begrenzung der Bearbeitungszeit auf 10 Sekunden pro Item führt sowohl direkt, also in der CV, als auch indirekt, durch die PBV zu T2, zu starken Streuungen der Mittelwerte der jeweiligen Bedingungskonstellationen. Die Anzeige der Gesamtbearbeitungsdauer hat nur einen sehr geringen Effekt in diese Richtung.

Allgemein kann festgestellt werden, dass sich die Effekte, die sich als signifikant erweisen, nicht immer für alle Subtest generalisieren lassen. Dies kann nun daran, dass die Effekte zufällig so groß sind, oder an den Subtesteigenschaften liegen. Hinsichtlich der Präsentation ist es nachvollziehbar, dass die Gesamtansicht aller Items das Zählen jeder achten Ziffer (erst 0, dann 1 usw.; Subtest 9) erleichtert - insbesondere wenn der Subtest zum ersten Mal bearbeitet wird (vgl. Abbildung 10.11). Vor allem die Notwendigkeit, einen „weiter“-Button zu klicken, ist bei Subtest 9 hinderlich (vgl. Abbildung 10.14). Für Subtest 10 führt die Isolierung der einzelnen Items zu einem Vorteil, da dies offenbar den Vergleich von linker und rechter Seite erleichtert (vgl. Abbildung 3.4, S. 36). Für Subtest 2 zeigt sich dieser Vorteil ebenso, allerdings nur für die Testung zu T1 (vgl. Abbildung 10.11). Die drei unterschiedlichen Arten der Antwortabgabe zeigen für sich betrachtet nur einen geringen Einfluss. Es fällt lediglich auf, dass in den Subtests 1, 9 und 10 die Bedingung, in der die Möglichkeit gegeben wird, über einen Klick zu bereits bearbeiteten Aufgaben zu gelangen, durchwegs zu den niedrigsten Ergebnissen führt (vgl. Abbildung 10.9). Wird jedoch der Bearbeitungszeitpunkt kontrolliert, fällt auf, dass dieser Effekt bei den Subtests 3, 5 und 8 nur in der CV zu T1 auftritt. Werden die Subtests zu T2 in der CV bearbeitet, resultiert häufig der beste Wert. Für die Subtests 9 und 10 scheint die Zeitersparnis, wenn sofort zum nächsten Item gewechselt wird, von besonderem Vorteil gewesen zu sein, insbesondere wenn keine Zeitvorgabe gemacht wurde (vgl. Abbildung 10.16).

Die Zeitvorgabe erweist sich ohnehin als der am meisten Varianz erzeugende Faktor, sofern die Anzahl statistisch signifikanter Effekte besondere Relevanz haben kann. Für sich genommen führt die Begrenzung auf 10 Sekunden pro Item bei den Subtests 3, 5 und 8 zu geringeren Werten (vgl. Abbildung 10.10). Auch bei Berücksichtigung des Bearbeitungszeitpunktes lässt sich dieser Effekt relativ zu den beiden anderen Bedingungen zu beiden Möglichkeiten für die Subtests 1, 3, 4 und 8 beobachten (vgl. Abbildung 10.13).

In Tabelle 10.24 werden die Ergebnisse der Untersuchung 2 hinsichtlich der Äquivalenz unter dem Aspekt der 15 verschiedenen Bedingungen, die sich aus den drei variierten Faktoren bilden, zusammengefasst. Die Tabelle kann ebenso quer bezüglich des jeweiligen Subtests gelesen werden. Für Subtest 1 zeigen sich über die Bedingungen hinweg vorwiegend mittlere Effekte, wobei in der PBV immer besser abgeschnitten wurde. Ein ähnliches Bild ergibt sich für die Subtest 2, 6 und 11; hier können einige Bedingungen als durchaus äquivalent bzgl. der Mittelwerte angesehen werden. Für Subtest 3 ergeben sich vorwiegend leicht negative Werte (PBV besser), allerdings auch einige Werte um Null und ein hoher positiver Wert. Bei Subtest 4 liegen die meisten Effektgrößen im sehr geringen Bereich. Bei der Einzelpräsentation der Items hat scheinbar die PBV leichte Vorteile; werden alle Items präsentiert, zeigen sich eher geringe, aber positive Werte (CV besser). Bei den Subtests 5, 7 und 8 kommen nur wenige Werte nahe Null vor. Die meisten Effekte können als mittel stark bezeichnet werden und sind meist negativ; insbesondere Subtest 5 weist jedoch auch einige kleine positive Effekte auf. Für Subtest 9 zeigen

sich sowohl starke positive (besonders wenn alle Items in der CV präsentiert werden) als auch starke negative Effekte; daneben kommen in einigen Bedingungen fast keine Unterschiede vor. Für Subtest 10 zeigen sich bei allen Bedingungen, insbesondere bei der Präsentation aller Items in der CV, extreme Vorteile für die PBV. Mit Ausnahme von Subtest 9 können die Mittelwerte der Effektstärken über die Bedingungen hinweg durchaus als Maßzahl für die Möglichkeit, äquivalente Bedingungen zu schaffen, gesehen werden.

Bei Betrachtung der Effektgrößen der Subtests in den einzelnen Bedingungen in Tabelle 10.24 fallen die stärksten Abweichungen hinsichtlich eines augenfälligen Musters in den Subtests 9 und 10 auf. Dies bestätigt sich in der Überprüfung der Homogenität der Effekte. Der kritische Wert für die χ^2 -verteilte Prüfgröße Q_i liegt bei Betrachtung aller Subtests bei $\chi^2(.90; 11 - 1) = 15.99$. Dieser Wert wird in zehn der 15 Bedingungen überschritten. In vier der fünf Bedingungen mit homogenen Effektstärken wurden die Subtests 9 und 10 wegen ihrer besonderen Beschaffenheit jedoch nicht durchgeführt. Es zeigen sich somit nur in einer Bedingung homogene Effekte bei Beachtung aller Subtests. Die gewichtete Effektgröße (unter Berücksichtigung der Interkorrelationen, vgl. Tabelle 11.1) hat in diesem Fall eine mittlere Höhe von -0.43 ($\hat{\sigma}_{ss}^{(5)} = 0.04$).

Tabelle 10.24: Effektstärken je Subtest je Bedingung sowie dazugehörige Mittelwerte und Parameter der multivariaten Betrachtung

Präsentation		Einzel									Alle							
Antwort-format		Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar			
Zeitvorgabe		Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ohne	Ge-samt	Ohne	Ge-samt		
Bedingung		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	M_{d_p}	SD_{d_p}
Subtest	1	-0.32	-0.28	-0.30	-0.46	-0.18	-0.57	-1.33	-0.39	-0.74	-0.46	-0.37	-0.53	-0.36	-0.72	-0.63	-0.51	0.28
	2	-0.28	-0.40	-0.06	-0.59	-0.11	-0.28	-0.09	0.06	-0.20	-0.32	-0.25	-0.51	-0.68	-0.50	0.00	-0.28	0.22
	3	-0.20	0.71	-0.11	-0.14	-0.35	-0.15	-0.03	0.11	-0.56	-1.11	0.00	-0.14	-0.14	-0.34	-0.47	-0.20	0.38
	4	0.05	0.04	-0.25	0.00	-0.18	-0.10	-0.05	-0.02	-0.83	0.50	0.13	0.24	0.00	0.00	0.04	-0.03	0.28
	5	-0.28	-0.07	-0.40	-0.51	0.27	-0.58	-0.40	0.20	-0.25	-0.18	0.22	-0.77	-0.21	-0.15	0.16	-0.20	0.31
	6	-0.01	-0.44	-0.66	-0.24	-1.13	-0.68	-0.68	-0.55	-0.82	-0.56	-0.44	-0.66	-0.33	-0.21	-0.23	-0.51	0.28
	7	-0.19	-0.42	-0.38	-0.69	-0.38	0.11	-0.99	-0.68	-0.78	-0.21	-0.06	-0.41	-0.62	-0.74	-0.17	-0.44	0.31
	8	0.13	-0.08	-0.33	-0.90	-0.36	-0.91	-0.22	-0.35	-0.71	-0.24	-0.29	-0.27	-0.25	-0.55	-0.62	-0.40	0.29
	9	1.54	0.05	-	-0.16	-0.25	-	-1.49	-1.01	-	-	-	1.58	0.76	1.27	1.26	0.35	1.09
	10	-0.52	-4.95	-	-1.60	-0.63	-	-1.48	-0.57	-	-	-	-4.45	-4.12	-4.95	-5.03	-2.83	2.02
	11	-0.39	-0.36	0.04	0.06	-0.30	-0.08	-0.93	-0.10	-0.43	-0.34	-0.30	-0.28	0.10	-0.63	-0.09	-0.27	0.27
Parameter	M_{d_i}	-0.04	-0.56	-0.27	-0.48	-0.33	-0.36	-0.70	-0.30	-0.59	-0.32	-0.15	-0.56	-0.53	-0.68	-0.53	-0.48	0.19
	SD_{d_i}	0.56	1.49	0.21	0.48	0.35	0.34	0.58	0.38	0.25	0.42	0.23	1.44	1.25	1.52	1.58	0.82	
	Q_i	28.49	72.10	4.45	25.74	13.91	13.16	41.70	16.52	7.14	21.9	5.06	95.0	128.1	83.4	137.4		
	$\hat{\delta}_i$	0.02	0.26	-0.27	-0.37	-0.43	-0.33	-1.07	-0.48	-0.68	-0.25	-0.21	0.60	0.12	0.47	0.16		
	$\hat{\sigma}_{ss}^{(i)}$	0.04	0.05	0.05	0.05	0.04	0.05	0.04	0.04	0.05	0.05	0.05	0.04	0.04	0.04	0.05		
	$u_{\hat{\delta}_{ii}}$	-0.08	0.16	-0.39	-0.48	-0.52	-0.44	-1.17	-0.58	-0.79	-0.36	-0.33	0.50	0.02	0.37	0.06		
	$o_{\hat{\delta}_{ii}}$	0.13	0.37	-0.16	-0.26	-0.33	-0.22	-0.97	-0.37	-0.56	-0.14	-0.10	0.70	0.22	0.57	0.27		
	\bar{n}_i	16.45	16.09	16.00	15.91	18.82	16.78	18.64	17.45	16.22	15.7	15.78	17.0	18.55	17.2	17.55		
Parameter \ {9;10}	M_{d_i}	-0.17	-0.14	-0.27	-0.39	-0.30	-0.36	-0.52	-0.19	-0.59	-0.32	-0.15	-0.37	-0.28	-0.43	-0.22	-0.31	0.13
	SD_{d_i}	0.18	0.36	0.21	0.32	0.37	0.34	0.48	0.31	0.25	0.42	0.23	0.30	0.26	0.26	0.29	0.16	
	Q_i	2.82	10.96	4.45	9.43	13.11	13.16	21.31	8.88	7.14	21.9	5.06	10.3	6.39	6.35	8.35		
	$\hat{\delta}_i$	-0.15	-0.15	-0.27	-0.32	-0.37	-0.33	-0.59	-0.25	-0.68	-0.25	-0.21	-0.31	-0.24	-0.45	-0.34		
	$\hat{\sigma}_{ss}^{(i)}$	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05		
	$u_{\hat{\delta}_{ii}}$	-0.27	-0.27	-0.40	-0.44	-0.49	-0.45	-0.71	-0.37	-0.80	-0.38	-0.34	-0.44	-0.36	-0.58	-0.46		
	$o_{\hat{\delta}_{ii}}$	-0.03	-0.02	-0.15	-0.19	-0.24	-0.21	-0.46	-0.12	-0.55	-0.13	-0.09	-0.19	-0.11	-0.33	-0.22		
	\bar{n}_i	16.22	16.22	16.00	16.11	16.67	16.78	16.33	16.22	16.22	15.7	15.78	15.7	15.89	15.8	16.11		

Bemerkungen: Da $d_{il} = c(\bar{Y}_{il}^{CV} - \bar{Y}_{il}^{PBV})/s_{il}^{(i)}$ (vgl. Formel (7-5)), bedeutet eine negative Effektgröße, dass in der PBV die höhere Leistung erzielt wurde. $d_{il} < .15$ sind grau markiert. Die Homogenität der Effektgrößen muss für $Q_i > \chi^2(.90; 11 - 1) = 15.99$ bzw. $Q_i > \chi^2(.90; 9 - 1) = 13.36$ abgelehnt werden. Q_i kleiner dieser Marken sind grau hinterlegt. $u_{\hat{\delta}_{ii}}$ - Untergrenze, $o_{\hat{\delta}_{ii}}$ - Obergrenze des 90 %-Konfidenzintervall von $\hat{\delta}_i$. Teilweise wurden für eine bessere Darstellung Werte auf eine Nachkommastelle gerundet.

Wegen der im Vergleich zu den übrigen Subtests auffälligen Ergebnisse wurden dieselben Berechnungen hinsichtlich der Äquivalenz ohne die beiden Subtests 9 und 10 berechnet. In dieser Konstellation kann die Homogenität der Effekte in 13 der 15 Bedingungen angenommen werden. Die gewichteten Effekte bewegen sich im Bereich zwischen -0.15 und -0.68, wobei zwei Effekte als gering ($|\hat{\delta}_i| < 0.20$) und die Mehrheit als mittel ($|\hat{\delta}_i| \in [0.20; 0.50]$) zu werten sind. Kein Konfidenzintervall der homogenen, gewichteten Effekte beinhaltet die Null (selbst wenn die Konfidenzgrenze auf $\alpha = .05$ gesetzt werden würde).

Tabelle 10.25: Effektstärken je Subtest je Bedingung sowie dazugehörige Mittelwerte und Parameter der multivariaten Betrachtung für die Gruppen, die die CV zu T1 und die PBV zu T2 bearbeiteten

Präsentation		Einzel									Alle							
Antwort-format		Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar			
Zeitvorgabe		Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ohne	Ge-samt	Ohne	Ge-samt		
Bedingung		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
Subtest	1	-3.27	-2.10	-0.99	-2.85	-2.77	-1.70	-6.27	-0.97	-1.98	-1.69	-1.10	-0.82	-0.62	-1.57	-1.53	-1.69	-1.10
	2	-1.06	-1.32	-2.01	-2.60	-0.51	-2.18	-1.06	-0.59	-0.50	-1.04	-1.20	-3.07	-2.71	-1.04	-0.71	-1.04	-1.20
	3	-1.12	-1.14	-0.24	-1.06	-1.55	-0.67	-1.48	-0.83	-1.67	-2.73	-1.26	-1.49	-1.14	-1.03	-1.27	-2.73	-1.26
	4	-0.14	-0.65	-0.69	-0.62	-0.71	-0.15	-0.22	-0.41	-1.32	0.14	-0.68	0.00	-0.12	-0.54	-0.83	0.14	-0.68
	5	-0.79	-0.31	-0.80	-1.06	-0.50	-0.45	-0.67	0.17	-0.98	-0.61	0.34	-1.01	-1.88	-0.67	-0.09	-0.61	0.34
	6	-1.43	-0.90	-1.13	-0.83	-2.59	-1.35	-1.47	-0.98	-1.84	-2.35	-2.21	-1.27	-0.46	-1.11	-0.63	-2.35	-2.21
	7	-0.94	-1.15	-0.82	-1.09	-0.67	-0.33	-1.61	-1.60	-1.41	-0.36	-0.85	-0.92	-1.06	-2.39	-0.48	-0.36	-0.85
	8	-0.95	-1.09	-0.54	-1.85	-0.91	-1.76	-1.01	-1.15	-0.89	-0.48	-1.27	-0.55	-0.67	-1.16	-1.49	-0.48	-1.27
	9	1.16	-0.45	-	-1.10	-0.67	-	-2.58	-2.24	-	-	-	1.17	0.17	1.38	1.35	-	-
	10	-0.49	-15.8	-	-1.40	-0.83	-	-1.61	-1.26	-	-	-	-5.33	-4.66	-5.78	-6.09	-	-
	11	-0.78	-0.51	-0.29	-1.39	-1.77	-0.13	-1.00	-0.66	-0.68	-1.30	-1.00	-0.38	-0.44	-1.77	-0.64	-1.30	-1.00
Parameter	M_{d_i}	-0.89	-2.32	-0.84	-1.44	-1.23	-0.97	-1.73	-0.96	-1.25	-1.16	-1.02	-1.24	-1.23	-1.42	-1.13	-1.16	-1.02
	SD_{d_i}	1.05	4.51	0.53	0.71	0.83	0.78	1.63	0.63	0.52	0.95	0.67	1.70	1.40	1.72	1.83	0.95	0.67
	Q_i	42.65	39.23	10.70	17.86	34.50	25.02	58.76	20.67	14.35	41.2	27.03	65.6	80.15	64.0	77.65		
	$\hat{\delta}_i$	-0.49	-0.17	-0.54	-1.01	-0.68	-0.65	-1.24	-1.05	-1.21	-0.57	-0.83	0.37	0.04	0.12	-0.23		
	$\hat{\sigma}_{ss}^{(i)}$	0.09	0.08	0.09	0.10	0.07	0.09	0.09	0.09	0.12	0.11	0.11	0.07	0.07	0.08	0.09		
	$u_{\hat{\delta}_{ij}}$	-0.64	-0.31	-0.69	-1.16	-0.81	-0.80	-1.39	-1.20	-1.38	-0.73	-1.00	0.24	-0.09	-0.02	-0.37		
	$o_{\hat{\delta}_{ij}}$	-0.34	-0.03	-0.40	-0.85	-0.55	-0.50	-1.09	-0.91	-1.04	-0.40	-0.66	0.51	0.16	0.26	-0.08		
	\bar{n}_i	9.36	9.27	9.22	8.82	10.45	9.11	10.18	9.18	8.44	8.22	8.44	9.36	10.18	9.64	10.00		
Parameter \ {9;10}	M_{d_i}	-1.16	-1.02	-0.84	-1.48	-1.33	-0.97	-1.64	-0.78	-1.25	-1.16	-1.02	-1.06	-1.01	-1.25	-0.85	-1.12	0.24
	SD_{d_i}	0.86	0.52	0.53	0.79	0.88	0.78	1.79	0.50	0.52	0.95	0.67	0.88	0.82	0.57	0.49	0.47	
	Q_i	23.31	11.94	10.70	17.79	32.22	25.02	44.08	12.23	14.35	41.2	27.03	29.8	25.14	13.5	14.79		
	$\hat{\delta}_i$	-0.96	-0.87	-0.54	-1.03	-0.95	-0.65	-0.78	-0.82	-1.21	-0.57	-0.83	-0.63	-0.22	-1.10	-0.97		
	$\hat{\sigma}_{ss}^{(i)}$	0.10	0.10	0.09	0.11	0.11	0.09	0.11	0.10	0.12	0.11	0.11	0.10	0.08	0.11	0.09		
	$u_{\hat{\delta}_{ij}}$	-1.13	-1.05	-0.71	-1.21	-1.13	-0.82	-0.96	-1.00	-1.39	-0.75	-1.02	-0.80	-0.38	-1.28	-1.14		
	$o_{\hat{\delta}_{ij}}$	-0.78	-0.70	-0.38	-0.85	-0.77	-0.48	-0.60	-0.65	-1.02	-0.39	-0.65	-0.46	-0.06	-0.92	-0.80		
	\bar{n}_i	9.33	9.44	9.22	9.11	9.22	9.11	9.00	8.56	8.44	8.22	8.44	8.56	8.89	9.22	9.44		

Bemerkungen: Da $d_{il} = c(\bar{Y}_{il}^{CV} - \bar{Y}_{il}^{PBV})/s_{il}^{(i)}$ (vgl. Formel (7-5)), bedeutet eine negative Effektgröße, dass in der PBV die höhere Leistung erzielt wurde. $d_{il} < .15$ sind grau markiert. Die Homogenität der Effektgrößen muss für $Q_i > \chi^2(.90; 11 - 1) = 15.99$ bzw. $Q_i > \chi^2(.90; 9 - 1) = 13.36$ abgelehnt werden. Q_i kleiner dieser Marken sind grau hinterlegt. $u_{\hat{\delta}_{ii}}$ - Untergrenze, $o_{\hat{\delta}_{ii}}$ - Obergrenze des 90 %-Konfidenzintervall von $\hat{\delta}_i$. Teilweise wurden für eine bessere Darstellung Werte auf eine Nachkommastelle gerundet.

Um die Einflüsse von Modus und zweiter Testung verdeutlichen und unterscheiden zu können, wurden in Tabelle 10.25 dieselben Berechnungen nur für die Gruppe der Versuchspersonen durchgeführt, die die CV zu T1 bearbeiteten, und in Tabelle 10.26 für diejenigen, die die CV zu T2 durchführten. Die Q_i -Werte zeigen, dass es durch die größere Stichprobe und die Beachtung beider Messzeitpunkte bei der Gesamtbetrachtung zu Homogenisierungseffekten kommt. Wird die CV zu T1 durchgeführt, resultieren die heterogensten Effekte. Insgesamt wird anhand der Effektwerte deutlich, was auch schon insbesondere in Abbildung 10.7 zu beobachten war: Der Vorteil einer zweiten Testung ist in den meisten Fällen größer, wenn zuerst die CV und dann die PBV bearbeitet wurde.

Wird die CV zu T1 durchgeführt, zeigt nur Subtest 4 in mehreren Bedingungen geringe Effektstärken. Für Subtest 9 sind als einziger Test die Effektstärken bei der Präsentation aller Items positiv, d. h. die Ergebnisse sind bei der ersten Testung als CV im Mittel besser als bei der zweiten Testung als PBV; dies ist zudem auch in Bedingung 1 der Fall. Die gewichteten Effekte der statistisch als homogen annehmbaren Bedingungen weisen mittlere bis hohe Ausprägungen auf ($\hat{\delta}_2 = -0.87$, $\hat{\delta}_3 = -0.54$, $\hat{\delta}_8 = -0.82$). Kein Konfidenzintervall beinhaltet die Null.

Tabelle 10.26: Effektstärken je Subtest je Bedingung sowie dazugehörige Mittelwerte und Parameter der multivariaten Betrachtung für die Gruppen, die die PBV zu T1 und die CV zu T2 bearbeiteten

Präsentation		Einzel									Alle							
Antwort-format		Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar			
Zeitvorgabe		Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ohne	Ge-samt	Ohne	Ge-samt		
Bedingung		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	M_{d_p}	SD_{d_p}
Subtest	1	1.12	0.51	0.14	0.61	0.65	0.09	-0.30	0.31	-0.03	0.90	0.30	0.54	0.42	0.55	0.28	0.41	0.36
	2	0.68	1.05	0.87	0.60	0.55	0.64	0.25	1.70	0.58	0.26	0.69	2.35	1.30	0.74	1.48	0.92	0.57
	3	1.70	1.99	-0.02	0.89	0.59	0.53	1.94	1.58	0.65	0.36	1.12	1.18	1.93	2.25	2.13	1.25	0.73
	4	0.43	1.69	1.12	0.60	1.07	-0.05	0.23	0.49	-0.49	1.28	1.31	0.55	0.19	0.89	1.98	0.75	0.67
	5	0.43	0.35	0.19	0.29	1.71	-0.67	0.10	0.25	0.71	0.44	0.00	-0.43	0.58	1.29	0.61	0.39	0.59
	6	1.44	0.20	0.03	0.13	-0.23	-0.10	-0.11	0.12	0.11	0.20	0.46	0.60	-0.14	0.68	0.32	0.25	0.42
	7	0.99	1.11	0.24	0.73	0.50	2.42	0.60	0.28	-0.20	0.00	0.62	0.29	0.12	1.44	0.28	0.63	0.66
	8	1.82	1.46	-0.04	0.78	0.46	0.35	0.56	2.15	-0.45	0.24	1.34	0.34	0.60	0.75	1.64	0.80	0.73
	9	1.75	0.88	-	0.45	0.35	-	-0.85	-0.21	-	-	-	2.19	1.67	1.09	1.06	0.84	0.93
	10	-0.56	-2.93	-	-1.76	-0.17	-	-1.35	0.10	-	-	-	-3.90	-3.50	-3.68	-3.79	-2.15	1.59
	11	0.11	-0.09	0.28	0.45	0.53	0.00	-0.69	0.17	-0.08	0.34	0.26	-0.22	0.58	-0.04	0.52	0.14	0.34
Parameter	M_{d_i}	0.90	0.57	0.31	0.34	0.55	0.36	0.04	0.63	0.09	0.45	0.68	0.32	0.34	0.54	0.59	0.38	0.23
	SD_{d_i}	0.76	1.33	0.41	0.73	0.53	0.87	0.88	0.79	0.46	0.40	0.48	1.65	1.43	1.51	1.61	0.90	
	Q_i	29.31	53.92	7.41	27.04	12.10	22.70	35.14	28.10	12.22	7.98	9.61	80.1	89.11	47.4	102.2		
	$\hat{\delta}_i$	0.83	0.81	0.16	0.42	0.17	0.08	-0.51	0.08	-0.15	0.47	0.62	1.07	0.44	0.91	0.22		
	$\hat{\sigma}_{ss}^{(i)}$	0.11	0.12	0.12	0.11	0.08	0.11	0.08	0.09	0.10	0.11	0.12	0.10	0.09	0.09	0.11		
	$u_{\hat{\delta}_{ii}}$	0.67	0.64	-0.01	0.25	0.03	-0.08	-0.65	-0.07	-0.31	0.30	0.45	0.91	0.29	0.76	0.06		
	$o_{\hat{\delta}_{ii}}$	1.00	0.98	0.33	0.58	0.31	0.24	-0.37	0.23	0.01	0.64	0.79	1.22	0.59	1.06	0.38		
	\bar{n}_i	7.09	6.82	6.78	7.09	8.36	7.67	8.45	8.27	7.78	7.44	7.33	7.64	8.36	7.55	7.55		
Parameter \ {9;10}	M_{d_i}	0.97	0.92	0.31	0.56	0.65	0.36	0.29	0.78	0.09	0.45	0.68	0.58	0.62	0.95	1.03	0.61	0.28
	SD_{d_i}	0.60	0.72	0.41	0.24	0.52	0.87	0.74	0.79	0.46	0.40	0.48	0.81	0.64	0.64	0.77	0.35	
	Q_i	12.11	15.50	7.41	2.33	9.58	22.70	16.15	23.82	12.22	7.98	9.61	23.5	12.97	8.27	16.31		
	$\hat{\delta}_i$	0.94	0.76	0.16	0.59	0.39	0.08	0.18	0.43	-0.15	0.47	0.62	0.38	0.19	0.65	0.47		
	$\hat{\sigma}_{ss}^{(i)}$	0.12	0.13	0.12	0.12	0.11	0.11	0.11	0.12	0.10	0.11	0.12	0.11	0.11	0.13	0.12		
	$u_{\hat{\delta}_{ii}}$	0.75	0.56	-0.03	0.41	0.21	-0.10	-0.01	0.24	-0.33	0.29	0.43	0.20	0.01	0.46	0.27		
	$o_{\hat{\delta}_{ii}}$	1.14	0.96	0.35	0.78	0.57	0.26	0.37	0.62	0.02	0.65	0.81	0.56	0.38	0.85	0.66		
	\bar{n}_i	6.89	6.78	6.78	7.00	7.44	7.67	7.33	7.67	7.78	7.44	7.33	7.11	7.00	6.56	6.67		

Bemerkungen: Da $d_{il} = c(\bar{Y}_{il}^{CV} - \bar{Y}_{il}^{PBV})/s_{il}^{(i)}$ (vgl. Formel (7-5)), bedeutet eine negative Effektgröße, dass in der PBV die höhere Leistung erzielt wurde. $d_{il} < .15$ sind grau markiert. Die Homogenität der Effektgrößen muss für $Q_i > \chi^2(.90; 11 - 1) = 15.99$ bzw. $Q_i > \chi^2(.90; 9 - 1) = 13.36$ abgelehnt werden. Q_i kleiner dieser Marken sind grau hinterlegt. $u_{\hat{\delta}_{ii}}$ - Untergrenze, $o_{\hat{\delta}_{ii}}$ - Obergrenze des 90 %-Konfidenzintervall von $\hat{\delta}_i$. Teilweise wurden für eine bessere Darstellung Werte auf eine Nachkommastelle gerundet.

Die Effekte für die Gruppen, die die CV zu T2 bearbeiteten, sind deutlich geringer. Nur bei Subtest 10 sind die Ergebnisse der PBV mit einer Ausnahme (Bedingung 8) konsistent besser. Die anderen Bedingungskonstellationen mit negativen Effektstärken (20 von 145 Vergleichen, also

rd. 14 %) treten gehäuft auf, wenn das Antwortformat korrigierbar war; zudem ist eine Häufung bei der Zeitbeschränkung von 10 Sekunden pro Item zu beobachten. In Bedingung 8 (Einzeltrepräsentation, korrigierbar, 10s Beschränkung) resultiert auch der einzige negative gewichtete mittlere Effekt bei Beibehaltung der Homogenitätshypothese ($\hat{\delta}_9 = -0.15$), wobei die Subtests 2, 3, 5 und 6 positive Effekte aufweisen. Insgesamt zeigen mehrere Bedingungen homogene Effekte über die Subtests hinweg, wenn die CV zu T2 bearbeitet wird. Neben dem negativen Wert reicht das Spektrum von 0.16 bis 0.94; die einzige homogene Bedingung für alle elf Subtests ist Bedingung 5; dabei ergibt sich ein gewichteter Effekt von $\hat{\delta}_5 = 0.17$. Werden die Subtests 9 und 10 außen vor gelassen, erhöht sich dieser Wert jedoch auf 0.39. Ohne die beiden Subtests sind die Effekte für Bedingung 3 ($\hat{\delta}_3 = 0.16$) und Bedingung 9 ($\hat{\delta}_9 = -0.15$) statistisch nicht von Null verschieden.

Tabelle 10.27: Effektstärken je Subtest und Bedingung sowie dazugehörige Mittelwerte und Parameter der multivariaten Betrachtung des Vergleichs PBV vs. CV zu T1

Präsentation		Einzeln									Alle							
Antwort-format	Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar				
	Zeitvorgabe	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ohne	Ge-samt	Ohne	Ge-samt		
	Bedingung	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
Subtest	1	-6.55	-1.16	-0.66	-1.35	-4.93	-2.39	-5.61	-0.91	-1.90	-1.47	-1.18	-0.53	0.42	-1.74	-1.14	-2.07	2.00
	2	-0.48	-2.28	-0.17	-1.89	-0.03	-0.97	-2.11	-0.57	-0.36	-0.98	-1.11	-2.11	-1.04	-0.54	-0.88	-1.03	0.74
	3	-0.71	2.97	-0.28	0.03	0.44	-0.73	-0.12	-0.95	-1.12	0.57	0.52	-0.84	-0.58	0.32	-0.47	-0.06	1.01
	4	0.04	-0.34	0.46	-0.63	-0.49	1.13	-0.21	-0.41	-1.68	-0.13	0.38	0.58	-0.59	-0.69	0.34	-0.15	0.67
	5	-1.04	0.26	-0.91	0.46	-1.47	0.21	-0.34	0.39	0.00	-0.35	-0.71	0.21	-0.42	-0.72	-0.16	-0.31	0.58
	6	-0.17	-1.49	-0.20	-2.03	-1.61	-1.66	-0.84	-0.53	-0.79	-0.72	-1.45	-0.95	-0.53	-0.60	-0.68	-0.95	0.56
	7	-0.63	-1.45	-1.34	-0.64	-0.32	-0.05	-0.83	-0.48	-1.28	-0.09	-0.10	-0.86	-0.91	-0.40	0.66	-0.58	0.57
	8	-0.17	-0.39	-0.88	-1.04	-0.19	-0.93	-0.01	-0.19	-0.81	-0.54	0.49	-0.26	-0.60	-1.16	-0.66	-0.49	0.45
	9	1.91	-0.54	-	-1.49	-0.75	-	-2.69	-2.08	-	-	-	1.39	0.53	1.55	1.45	-0.07	1.66
	10	0.03	-12.4	-	-2.30	-0.49	-	-1.96	-0.97	-	-	-	-5.01	-4.24	-4.52	-4.29	-3.61	3.57
	11	-0.01	-0.33	-0.15	-0.13	-2.18	-0.76	-0.70	0.00	-0.39	-0.85	-2.23	-1.02	-1.01	0.44	-1.67	-0.73	0.79
Parameter	M_{d_i}	-0.71	-1.55	-0.46	-1.00	-1.09	-0.68	-1.40	-0.61	-0.93	-0.51	-0.60	-0.85	-0.81	-0.73	-0.68	-0.91	0.31
	SD_{d_i}	2.08	3.83	0.54	0.90	1.49	1.03	1.66	0.64	0.63	0.59	0.98	1.66	1.25	1.53	1.48	1.07	
	Q_i	47.44	82.88	10.14	34.98	48.63	36.19	56.88	22.03	24.31	10.2	22.77	55.7	53.69	66.2	69.31		
	$\hat{\delta}_i$	0.42	-0.06	-0.32	-0.87	-0.21	-0.77	-1.11	-0.74	-0.93	-0.60	-0.39	0.34	-0.01	0.31	-0.23		
	$\hat{\sigma}_{ss}^{(i)}$	0.10	0.10	0.12	0.12	0.08	0.13	0.10	0.09	0.12	0.12	0.12	0.09	0.08	0.10	0.12		
	$u_{\hat{\delta}_{ii}}$	0.26	-0.21	-0.49	-1.04	-0.35	-0.94	-1.27	-0.89	-1.10	-0.77	-0.56	0.19	-0.15	0.16	-0.40		
	$o_{\hat{\delta}_{ii}}$	0.58	0.10	-0.14	-0.70	-0.06	-0.59	-0.95	-0.59	-0.76	-0.43	-0.21	0.49	0.14	0.47	-0.06		
	\bar{n}_i	7.09	6.82	5.55	7.09	8.36	6.27	8.45	8.27	6.36	7.44	7.33	7.64	8.36	7.55	7.55		
Parameter \ {9;10}	M_{d_i}	-1.08	-0.47	-0.46	-0.80	-1.20	-0.68	-1.20	-0.40	-0.93	-0.51	-0.60	-0.64	-0.58	-0.57	-0.52	-0.71	0.27
	SD_{d_i}	2.08	1.51	0.54	0.85	1.64	1.03	1.77	0.42	0.63	0.59	0.98	0.78	0.44	0.67	0.72	0.61	
	Q_i	31.94	46.49	10.14	25.27	48.19	36.19	33.57	10.95	24.31	10.2	22.77	24.4	7.41	13.8	19.43		
	$\hat{\delta}_i$	0.09	-0.71	-0.32	-0.68	-0.22	-0.77	-0.40	-0.48	-0.93	-0.60	-0.39	-0.47	-0.53	-0.78	-0.53		
	$\hat{\sigma}_{ss}^{(i)}$	0.11	0.12	0.12	0.12	0.13	0.13	0.13	0.11	0.12	0.12	0.12	0.11	0.12	0.13	0.13		
	$u_{\hat{\delta}_{ii}}$	-0.10	-0.91	-0.51	-0.87	-0.42	-0.96	-0.59	-0.67	-1.12	-0.79	-0.58	-0.65	-0.72	-0.98	-0.72		
	$o_{\hat{\delta}_{ii}}$	0.27	-0.52	-0.13	-0.49	-0.03	-0.57	-0.20	-0.30	-0.74	-0.42	-0.19	-0.28	-0.34	-0.58	-0.33		
	\bar{n}_i	6.89	6.78	6.78	7.00	7.44	7.67	7.33	7.67	7.78	7.44	7.33	7.11	7.00	6.56	6.67		

Bemerkungen: Da $d_{il} = c(\bar{Y}_{il}^{CV} - \bar{Y}_{il}^{PBV})/s_{il}^{(i)}$ (vgl. Formel (7-5)), bedeutet eine negative Effektgröße, dass in der PBV die höhere Leistung erzielt wurde. $d_{il} < .15$ sind grau markiert. Die Homogenität der Effektgrößen muss für $Q_i > \chi^2(.90; 11 - 1) = 15.99$ bzw. $Q_i > \chi^2(.90; 9 - 1) = 13.36$ abgelehnt werden. Q_i kleiner dieser Marken sind grau hinterlegt. $u_{\hat{\delta}_{ii}}$ - Untergrenze, $o_{\hat{\delta}_{ii}}$ - Obergrenze des 90 %-Konfidenzintervall von $\hat{\delta}_i$. Teilweise wurden für eine bessere Darstellung Werte auf eine Nachkommastelle gerundet.

In der Gesamtbetrachtung (Tabelle 10.24) wird deutlich, dass der leistungshemmende Effekt der CV stärker ist als der leistungsfördernde Effekt der zweiten Testung. Sowohl die eins-zu-eins Abbildung der PBV (Bedingung 14) als auch die Bedingungen der meisten Übersetzungen bestehender Tests auf den Computer (Bedingungen 1, 2, 4 und 5) lassen nicht auf äquivalente Testungen schließen.

Die Overall-Parameter des Vergleichs der jeweils ersten Testung entweder als PBV oder als CV zeigen, dass die Effekte noch heterogener hinsichtlich der Subtests ausfallen als in den bisherigen Vergleichen. Dies ist natürlich zum einen auf das Setting zurückzuführen, da nun jeweils die Ergebnisse von zwei verschiedenen Stichproben verglichen werden. Dennoch lassen sich Ähnlichkeiten hinsichtlich des Vergleichs der CV zu T1 mit der PBV zu T2 ausmachen. In den meisten Fällen resultieren bessere Werte für die PBV. Die Ausnahme bildet auch hier wieder Subtest 9, wenn alle Items zugleich präsentiert werden. Vermutlich unterstützt die Möglichkeit, mit der Maus die Ziffern und Zeichen abzufahren, die Sorgfalt und das Arbeitstempo. Warum Bedingung 1 zu besseren Werten in der CV führt, kann so allerdings nicht erklärt werden. Als homogen über die Subtests hinweg können nur die Effekte in Bedingung 3, 8 und 10 gewertet werden, und auch nur, wenn die Subtest 9 und 10 außen vor bleiben. Für alle drei Bedingungen zeigen sich mittlere negative gewichtete Effekte. In allen drei Bedingungen gibt es nur je einen Subtest, bei dessen Vergleich die CV höhere Werte erzielt (vgl. Tabelle 10.27). Auch die Mehrheit der übrigen mittleren Effekte (unabhängig davon, ob gewichtet oder nicht) entspricht einer mittleren bis starken Ausprägung.

Beim Vergleich der Modi zu T2 führt in etwa der Hälfte der Fälle der PBV zu besseren Ergebnissen. In den Bedingungen 1, 2, 13 und 15 sind hingegen die Leistungen der meisten Subtests in der CV höher. Lediglich Subtest 10 hat konsequent negative Effekte (vgl. auch Abbildung 10.7). Die einzige bezüglich der Effekte der Subtests homogene Bedingung (9) weist ein $\hat{\delta}_9$ von -0.15 auf, dessen Konfidenzintervall die Null enthält.

Generell muss festgestellt werden, dass sich weder ein Subtest als konsistent äquivalent, sprich hinsichtlich der variierenden Bedingungen unbeeinflussbar, erwiesen hat, noch dass eine Bedingung identifiziert werden konnte, die Äquivalenz a priori erwarten lässt. Daher erscheint es notwendig, quasi die Interaktion Subtest \times Bedingung genauer zu betrachten, um so zu Empfehlungen hinsichtlich der Umsetzung von Tests am Computer zu gelangen, die über den Rat, eine eigene Normierungsstudie aufzulegen, hinausgeht.

Als Kriterium für äquivalente Mittelwerte soll $|d_{i,l}| < 0.15$ dienen. Diese Grenze wurde aus folgendem Grund gewählt: Im Kontext von Intelligenz- und Leistungstests ist es üblich, die Messwerte auf IQ-Werte zu normieren, also auf Mittelwert 100 und Standardabweichung 15. Ein d von 0.15 bedeutet einen Mittelwertsunterschied von 2.25 IQ-Punkten. Dies entspricht in etwa einem Drittel des Messfehlers bei einer angenommenen Reliabilität von .80, was in diesem Kontext zwar willkürlich, aber nicht zu hoch gewählt erscheint.

Tabelle 10.28: Effektstärken je Subtest und Bedingung sowie dazugehörige Mittelwerte und Parameter der multivariaten Betrachtung des Vergleichs PBV vs. CV zu T2

Präsentation		Einzel									Alle							
Antwort-format		Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar			
Zeitvorgabe		Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ohne	Ge-samt	Ohne	Ge-samt		
Bedingung		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	M_{d_p}	SD_{d_p}
Subtest	1	3.91	-0.01	-0.04	-0.04	1.50	0.65	-0.37	0.41	-0.16	0.75	0.56	0.12	-1.48	1.11	0.11	0.47	1.18
	2	0.35	3.75	-0.14	0.19	-0.15	0.02	0.64	1.94	0.48	0.15	0.65	1.72	0.22	-0.16	2.10	0.78	1.11
	3	1.34	0.28	0.01	-0.48	-0.79	0.66	0.08	1.66	0.26	-2.06	-0.46	0.57	1.43	-0.57	0.87	0.19	0.98
	4	0.27	1.41	-1.33	0.77	0.34	-1.43	0.23	0.49	-0.23	1.87	0.29	-0.21	0.92	1.34	0.58	0.35	0.91
	5	1.00	-0.27	0.47	-1.18	3.21	-1.07	-0.33	-0.14	-0.56	0.06	1.70	-1.92	-0.19	1.54	0.82	0.21	1.30
	6	0.69	1.35	-0.76	1.00	-0.65	0.29	-0.59	-0.38	-0.74	-0.70	-0.03	0.28	0.00	0.06	0.43	0.02	0.66
	7	0.58	2.32	1.22	-0.25	-0.25	1.94	-1.26	-0.99	-0.24	-0.40	-0.22	0.19	0.01	-0.34	-1.51	0.05	1.08
	8	0.44	0.51	0.39	-0.05	-0.12	-0.62	-0.04	0.11	-0.55	0.42	-1.36	-0.87	0.37	1.14	0.00	-0.01	0.63
	9	1.03	1.22	-	0.77	0.53	-	-0.71	-0.34	-	-	-	2.00	1.07	0.87	0.90	0.73	0.77
	10	-1.51	-3.77	-	-0.77	-0.93	-	-0.94	-0.15	-	-	-	-4.23	-3.93	-4.86	-5.45	-2.65	1.97
	11	-0.84	-0.30	0.18	-0.01	0.74	0.80	-1.28	-0.29	-0.43	0.03	1.43	0.59	1.07	-1.62	1.67	0.12	0.95
Parameter	M_{d_i}	0.66	0.59	0.00	0.00	0.31	0.14	-0.42	0.21	-0.24	0.01	0.28	-0.16	-0.05	-0.14	0.05	0.02	0.29
	SD_{d_i}	1.36	1.89	0.73	0.67	1.20	1.05	0.62	0.88	0.40	1.07	0.94	1.73	1.52	1.84	2.05	0.93	
	Q_i	76.46	125.3	28.81	36.70	69.06	52.05	24.98	45.58	10.13	59.5	35.73	109.4	120.2	80.0	142.5		
	$\hat{\delta}_i$	0.34	0.28	-0.10	0.29	-0.52	0.43	-0.65	-0.18	-0.15	0.35	0.06	1.38	0.19	0.76	0.87		
	$\hat{\sigma}_{ss}^{(i)}$	0.09	0.09	0.09	0.08	0.06	0.09	0.07	0.07	0.10	0.11	0.11	0.07	0.08	0.07	0.08		
	$u_{\hat{\delta}_{ii}}$	0.19	0.14	-0.25	0.15	-0.64	0.28	-0.78	-0.31	-0.31	0.19	-0.10	1.24	0.05	0.63	0.73		
	$o_{\hat{\delta}_{ii}}$	0.49	0.43	0.05	0.44	-0.39	0.58	-0.53	-0.04	0.00	0.51	0.23	1.51	0.33	0.89	1.01		
	\bar{n}_i	9.36	9.27	9.22	8.82	10.45	9.11	10.18	9.18	8.44	8.22	8.44	9.36	10.18	9.64	10.00		
Parameter \ {9;10}	M_{d_i}	0.86	1.00	0.00	0.00	0.43	0.14	-0.33	0.31	-0.24	0.01	0.28	0.05	0.26	0.28	0.56	0.24	0.37
	SD_{d_i}	1.29	1.35	0.73	0.64	1.26	1.05	0.65	0.95	0.40	1.07	0.94	1.01	0.85	1.07	1.04	0.26	
	Q_i	49.85	71.54	28.81	26.85	63.05	52.05	21.22	41.91	10.13	59.5	35.73	52.1	33.12	59.9	42.77		
	$\hat{\delta}_i$	0.45	-0.13	-0.10	0.27	-0.52	0.43	-0.42	0.12	-0.15	0.35	0.06	0.44	0.07	0.24	0.42		
	$\hat{\sigma}_{ss}^{(i)}$	0.10	0.09	0.09	0.09	0.09	0.09	0.09	0.10	0.10	0.11	0.11	0.09	0.09	0.10	0.09		
	$u_{\hat{\delta}_{ii}}$	0.27	-0.30	-0.27	0.11	-0.68	0.27	-0.58	-0.05	-0.32	0.17	-0.11	0.28	-0.10	0.07	0.25		
	$o_{\hat{\delta}_{ii}}$	0.62	0.04	0.06	0.44	-0.35	0.60	-0.25	0.29	0.02	0.53	0.24	0.61	0.23	0.41	0.58		
	\bar{n}_i	9.33	9.44	9.22	9.11	9.22	9.11	9.00	8.56	8.44	8.22	8.44	8.56	8.89	9.22	9.44		

Bemerkungen: Da $d_{il} = c(\bar{Y}_{il}^{CV} - \bar{Y}_{il}^{PBV})/s_{il}^{(i)}$ (vgl. Formel (7-5)), bedeute eine negative Effektgröße, dass in der PBV die höhere Leistung erzielt wurde. Die Homogenität der Effektgrößen muss für $Q_i > \chi^2(.90; 11 - 1) = 15.99$ bzw. $Q_i > \chi^2(.90; 9 - 1) = 13.36$ abgelehnt werden. Q_i kleiner dieser Marken sind grau hinterlegt. $u_{\hat{\delta}_{ii}}$ - Untergrenze, $o_{\hat{\delta}_{ii}}$ - Obergrenze des 90 %-Konfidenzintervall von $\hat{\delta}_i$. Teilweise wurden für eine bessere Darstellung Werte auf eine Nachkommastelle gerundet.

Eine Möglichkeit der Überprüfung der Äquivalenz ist es, PBV und CV zu T1 zu vergleichen, womit im Folgenden begonnen wird. Zusätzlich werden die Ergebnisse des Vergleichs der Modi zu T2 eingestreut, da die Lerneffekte ja unabhängig von der Bedingung auftreten müssten. Für Subtest 1 resultieren in allen außer Bedingung 13 in der PBV bessere Ergebnisse. Außer für Subtest 9, dessen Besonderheit bei der Präsentationsform „alle“ bereits erwähnt wurde, zeigt Bedingung 13 bei den übrigen Subtests Vorteile für die PBV. Kein Wert davon deutet jedoch auf Äquivalenz hin. Für Subtest 2 ist der Effekt in Bedingung 5 klein genug, für Subtest 3 der in Bedingung 4. Die einzige Bedingung mit zumindest drei äquivalenten Ergebnissen ist Bedingung 1. Zu T2 ist der geringste Effekt von Bedingung 1 jedoch mit 0.27 zu groß. Für Subtest 11 zeigen sich zu T1 vier äquivalente Bedingungen (1, 3, 4 und 8), für Subtest 7 drei (6, 10 und 11). Die Subtests 6 und 9 weisen wie Subtest 1 keine geringen Effekte zu T1 auf; Subtest 9 zu T2 ebenfalls nicht. Subtest 4

weist zu T1 zwei äquivalente Bedingungen auf (1 und 10); in beiden wird keine Zeit angezeigt und es ist nur einmaliges Klicken auf eine Lösung möglich. Für Subtest 5 führt Bedingung 9 als jeweils einzige zu einem Nulleffekt; dasselbe gilt für Subtest 8 in Bedingung 7. Zu T2 weisen mehrere Subtests mehrere äquivalente Bedingungen auf; eine Bedingung ist jedoch für maximal drei Subtests äquivalent.

Werden die Mittelwerte von PBV und CV unabhängig vom Testzeitpunkt betrachtet, quasi gemittelt über den Lerneffekt, weisen immerhin drei Bedingungen (2, 6 und 8) Äquivalenz für je vier Subtests auf. Subtest 4 zeigt in diesen drei und in sechs weiteren Bedingungen (1, 4, 7, 13, 14 und 15) nur geringe Mittelwertsunterschiede. Ebenso weisen die Subtests 2, 3 und 11 kleine Effekte über alle Bedingungskonstellationen hinweg auf.

Insgesamt betrachtet kann keine Generalisierung aus den vorgestellten Ergebnissen abgeleitet bzw. gerechtfertigt werden. Es kann jedoch festgehalten werden, dass die Testung als CV mit Einzelitempräsentation keine spezifischen Nachteile gegenüber der Präsentation aller Items in gleicher Weise wie in der PBV aufweist. Kann ein Item nur einmal angeklickt werden, zeigen sich für einige als mittelschwer eingeschätzte Subtests Vorteile gegenüber der korrigierbaren Version und der Bedingung, in der ein Weiter-Button geklickt werden muss. Hierfür können zwei Gründe angeführt werden. Aufgrund der fehlenden Fehlerkorrekturmöglichkeit entsteht einerseits die Notwendigkeit, sorgfältiger zu arbeiten, was zwangsläufig zu weniger Flüchtigkeitsfehlern führt. Zudem wird Zeit gespart, die für die Bearbeitung von mehr Items genutzt wird. Beide Fälle würden somit die leistungshemmenden Einflüsse der CV etwas kompensieren und somit zu eher äquivalenten Ergebnissen führen. Die beiden sprachlichen Subtests (1 und 2) zeigen diesen Effekt jedoch nur bedingt. Insbesondere für Subtest 1 resultieren in einzelnen Bedingungen Effekte, die bei Gesamtbetrachtung nur als stichprobenbedingt erklärbar sind.

10.7 Vergleich der Korrelationen von PBV und CV

Gleichheit der Mittelwerte stellt die eine Bedingung für Äquivalenz dar. Ein zweites Kriterium ist die gleiche Reliabilität. Tabelle 10.11 zeigt, dass die interne Konsistenz als Schätzung der Untergrenze der Reliabilität für die meisten Bedingungen der CV durchaus vergleichbar ist. Die dritte wichtige Voraussetzung, im Hinblick auf die Konstruktion äquivalenter Verfahren vielleicht sogar wichtigste, ist die Korrelation der Testungen in den beiden Modi. In Tabelle 10.29 werden die Korrelationen von PBV mit der CV unabhängig vom Testzeitpunkt angegeben. Es zeigt sich wiederum ein sehr heterogenes Bild. Mit Ausnahme von Subtest 9 weisen sowohl jeder Subtest als auch jede Bedingung mindestens eine Korrelation größer .70 auf. Andererseits liegen ebenso für jeden Subtest sowie jede Bedingung einige Korrelationen vor, die statistisch nicht von Null verschieden sind. Zudem gibt es sogar einige negative Korrelationen.

Tabelle 10.29: Korrelationen (p -Value, n) von PBV und CV in den 15 Bedingungen

Präsentation		α PBV n=273 Bedingung	Einzel									Alle					
Antwort-format			Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar	
Zeitvorgabe			Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ein-zeln	Ohne	Ge-samt	Ohne	Ge-samt	Ohne	Ge-samt
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Subtest	1	.92	.73 (.01, 18)	.46 (.09, 15)	.74 (.01, 17)	.29 (.27, 17)	.66 (.01, 16)	.40 (.11, 17)	.62 (.01, 19)	.78 (.01, 15)	.46 (.08, 15)	.47 (.04, 19)	.74 (.01, 15)	.54 (.05, 14)	.82 (.01, 16)	.58 (.01, 17)	.08 (.76, 16)
	2	.90	.43 (.10, 16)	.70 (.01, 18)	.61 (.01, 15)	.33 (.20, 17)	.59 (.01, 17)	.48 (.06, 16)	.76 (.01, 17)	.69 (.01, 19)	.78 (.01, 15)	.60 (.02, 15)	.73 (.01, 19)	.54 (.04, 15)	-.59 (.03, 13)	.65 (.01, 17)	.64 (.01, 17)
	3	.76	.53 (.03, 17)	-.26 (.33, 16)	.31 (.21, 18)	.45 (.09, 15)	.45 (.07, 17)	.62 (.01, 17)	.25 (.36, 16)	.45 (.07, 17)	.51 (.03, 19)	.53 (.04, 15)	.34 (.24, 14)	.72 (.01, 19)	-.11 (.69, 15)	.30 (.32, 13)	.08 (.76, 17)
	4	.65	.80 (.01, 17)	.56 (.02, 17)	.11 (.69, 16)	.17 (.50, 18)	.38 (.16, 15)	.85 (.01, 17)	.34 (.18, 17)	.73 (.01, 16)	.12 (.64, 17)	.65 (.01, 19)	.12 (.67, 15)	.64 (.01, 15)	.75 (.01, 19)	.66 (.01, 15)	.02 (.94, 14)
	5	.74	.46 (.10, 14)	.45 (.07, 17)	.44 (.08, 17)	.64 (.01, 16)	.57 (.01, 18)	.44 (.10, 15)	.64 (.01, 16)	.78 (.01, 17)	.54 (.03, 16)	.01 (.98, 16)	.80 (.01, 19)	.64 (.01, 15)	.43 (.11, 15)	.44 (.06, 19)	.65 (.01, 15)
	6	.93	.43 (.11, 15)	.78 (.01, 14)	.73 (.01, 17)	.86 (.01, 17)	.64 (.01, 16)	.51 (.03, 18)	.68 (.01, 15)	.78 (.01, 16)	.48 (.05, 17)	.47 (.07, 15)	.31 (.23, 17)	.31 (.19, 15)	.79 (.01, 15)	.32 (.25, 15)	.71 (.01, 19)
	7	.86	.32 (.18, 19)	.76 (.01, 14)	.75 (.01, 14)	.44 (.08, 17)	.07 (.78, 17)	.69 (.01, 16)	.28 (.26, 18)	.58 (.02, 15)	.57 (.02, 16)	.71 (.01, 13)	.77 (.01, 14)	.54 (.03, 16)	.86 (.01, 19)	.35 (.20, 15)	.77 (.01, 15)
	8	.90	.45 (.10, 15)	.52 (.02, 19)	.73 (.01, 14)	-.02 (.96, 14)	.70 (.01, 17)	.35 (.16, 17)	.71 (.01, 16)	.30 (.22, 18)	.77 (.01, 15)	.69 (.01, 16)	.29 (.26, 17)	.47 (.09, 14)	.45 (.07, 17)	.70 (.01, 19)	.16 (.56, 15)
	9	.80	.20 (.44, 17)	.41 (.13, 15)	-	.44 (.13, 13)	.32 (.10, 29)	-	.26 (.17, 30)	.17 (.57, 14)	-	-	-	.41 (.02, 32)	.52 (.01, 30)	.59 (.01, 33)	.25 (.32, 18)
	10	.93	.38 (.12, 18)	.29 (.27, 16)	-	.37 (.14, 17)	.28 (.14, 28)	-	.64 (.01, 28)	.28 (.12, 32)	-	-	-	.57 (.03, 14)	.75 (.01, 31)	.71 (.01, 14)	.45 (.01, 30)
	11	.91	.55 (.04, 15)	.83 (.01, 16)	.83 (.01, 16)	.67 (.01, 14)	.38 (.13, 17)	.48 (.05, 18)	.38 (.20, 13)	.80 (.01, 13)	.57 (.02, 16)	.63 (.02, 13)	.74 (.01, 12)	.64 (.01, 14)	.62 (.02, 14)	.92 (.01, 12)	.64 (.01, 17)
Q_i			9.49	20.91	13.50	16.86	9.17	8.68	12.55	18.01	7.93	7.24	15.36	4.47	41.88	13.43	17.76
z_{+i}			.55	.61	.72	.50	.49	.64	.58	.67	.62	.61	.67	.61	.70	.68	.49
r_{+i}			.50	.54	.62	.46	.45	.56	.52	.59	.55	.55	.59	.54	.60	.59	.45
$u_{r_{+i}}$.39	.44	.51	.35	.35	.45	.43	.49	.43	.43	.48	.44	.52	.50	.34
$o_{r_{+i}}$.60	.63	.70	.56	.55	.65	.61	.67	.65	.65	.68	.63	.68	.67	.55

Bemerkungen: Korrelationen größer .70 sind hellgrau markiert. Die Homogenität der Korrelationskoeffizienten muss für $Q_i > \chi^2(.90; 11 - 1) = 15.99$ abgelehnt werden. Q_i kleiner dieser Marken sind grau hinterlegt. z_{+i} – gewichtete mittlere fischertransformierte Korrelation. r_{+i} – gewichtete mittlere Korrelation. $u_{r_{+i}}$ – Untergrenze, $o_{r_{+i}}$ – Obergrenze des 90 %-Konfidenzintervall von δ_i . Teilweise wurden für eine bessere Darstellung Werte auf eine Nachkommastelle gerundet.

In zehn der 15 Bedingungen kann die Homogenitätshypothese beibehalten werden ($Q_i < 15.99$), so dass eine Mittelung der Korrelationskoeffizienten als gerechtfertigt erachtet werden kann. Die höchsten Korrelationen zeigen sich für die Bedingung 3 ($r_{+3} = .72$). Dabei sind die Reihenfolgen der Probanden in der PBV außer für die Subtests 3, 4 und 5 der CV durchaus ähnlich. Die übrigen gemittelten Korrelationen liegen zwischen .49 und .68. Insgesamt zeigen sich die Subtests 2 und 11 am wenigsten anfällig hinsichtlich Veränderungen der Leistungsreihenfolgen über die Bedingungen hinweg. Die Subtests 3, 9 und 10 scheinen am anfälligsten. Höhere Werte

zeigen sich bei den Subtests 9 und 10 nur bei Bedingungen, in denen alle Items gleichzeitig dargestellt werden.

Tabelle 10.30: Korrelationen (p -Value, n) von PBV und CV für die Ausprägungen der drei Faktoren Präsentation, Zeitvorgabe und Antwortformat sowie ohne Differenzierung der Bedingungen (PC)

Faktor	Stufen	α PBV n=273	PC	Präsentation		Zeitvorgabe			Antwortformat		
				Einzel	Alle	Ohne	Gesamt	Einzel	Einmalig	Weiter	Korr.
				1-9	10-15	1, 4, 7, 10, 12, 14	2, 5, 8, 11, 13, 15	3, 6, 9	1, 2, 3, 10, 11	4, 5, 6, 12, 13	7, 8, 9, 14, 15
Subtest	1	.92	.60 (<.01, 246)	.59 (<.01, 149)	.59 (<.01, 97)	.56 (<.01, 104)	.65 (<.01, 93)	.60 (<.01, 49)	.65 (<.01, 84)	.60 (<.01, 80)	.54 (<.01, 82)
	2	.90	.59 (<.01, 246)	.61 (<.01, 150)	.57 (<.01, 96)	.59 (<.01, 97)	.58 (<.01, 103)	.63 (<.01, 46)	.64 (<.01, 83)	.42 (<.01, 78)	.69 (<.01, 85)
	3	.76	.35 (<.01, 245)	.35 (<.01, 152)	.36 (<.01, 93)	.49 (<.01, 95)	.18 (.08, 96)	.44 (<.01, 54)	.20 (.07, 80)	.52 (<.01, 83)	.35 (<.01, 82)
	4	.65	.51 (<.01, 247)	.50 (<.01, 150)	.55 (<.01, 97)	.56 (<.01, 101)	.52 (<.01, 96)	.44 (<.01, 50)	.51 (<.01, 84)	.57 (<.01, 84)	.45 (<.01, 79)
	5	.74	.54 (<.01, 245)	.54 (<.01, 146)	.54 (<.01, 99)	.57 (<.01, 96)	.63 (<.01, 101)	.45 (<.01, 48)	.54 (<.01, 83)	.50 (<.01, 79)	.60 (<.01, 83)
	6	.93	.61 (<.01, 245)	.68 (<.01, 145)	.49 (<.01, 100)	.58 (<.01, 96)	.65 (<.01, 97)	.56 (<.01, 52)	.59 (<.01, 78)	.62 (<.01, 85)	.60 (<.01, 82)
	7	.86	.54 (<.01, 238)	.46 (<.01, 146)	.66 (<.01, 92)	.38 (<.01, 98)	.61 (<.01, 94)	.66 (<.01, 46)	.63 (<.01, 74)	.50 (<.01, 85)	.48 (<.01, 79)
	8	.90	.48 (<.01, 243)	.49 (<.01, 145)	.46 (<.01, 98)	.55 (<.01, 94)	.40 (<.01, 103)	.55 (<.01, 46)	.49 (<.01, 81)	.44 (<.01, 79)	.54 (<.01, 83)
	9	.80	.29 (<.01, 231)	.26 (<.01, 118)	.46 (<.01, 113)	.26 (<.01, 125)	.34 (<.01, 106)	-	.28 (.13, 32)	.36 (<.01, 104)	.27 (<.01, 95)
	10	.93	.28 (<.01, 228)	.29 (<.01, 139)	.63 (<.01, 89)	.42 (<.01, 91)	.21 (.01, 137)	-	.23 (.19, 34)	.33 (<.01, 90)	.27 (<.01, 104)
	11	.91	.66 (<.01, 220)	.65 (<.01, 138)	.67 (<.01, 82)	.62 (<.01, 81)	.69 (<.01, 89)	.66 (<.01, 50)	.76 (<.01, 72)	.57 (<.01, 77)	.63 (<.01, 71)
Q_i			66.33	47.89	16.82	20.84	57.65	6.42	36.18	14.29	31.18
z_{+i}			.56	.56	.62	.56	.55	.63	.61	.54	.55
r_{+i}			.51	.51	.55	.51	.50	.56	.55	.49	.50
$u_{r_{+i}}$.48	.48	.51	.47	.46	.50	.50	.45	.45
$o_{r_{+i}}$.53	.54	.58	.54	.54	.61	.59	.54	.54

Bemerkungen: Korrelationen größer .70 sind hellgrau markiert. Die Homogenität der Korrelationskoeffizienten muss für $Q_i > \chi^2(.90; 11 - 1) = 15.99$ abgelehnt werden. Q_i kleiner dieser Marken sind grau hinterlegt. z_{+i} - gewichtete mittlere fischertransformierte Korrelation. r_{+i} - gewichtete mittlere Korrelation. $u_{r_{+i}}$ - Untergrenze, $o_{r_{+i}}$ - Obergrenze des 90 %-Konfidenzintervall von δ_i . Teilweise wurden für eine bessere Darstellung Werte auf eine Nachkommastelle gerundet.

Durch die zusammenfassende Betrachtung der jeweiligen Faktorenausprägungen lassen sich gewisse Stabilisierungstendenzen der Zusammenhänge auf für Leistungstests niedrigem Niveau ausmachen (Tabelle 10.30). Die einzige Korrelation größer .70 liegt für Subtest 11 für das Antwortformat „einmalig“ vor; nicht signifikante Korrelationen ($p > .05$) zeigen sich hingegen nur vier, wobei drei für eben diese Faktorausprägung vorliegen (Subtests 3, 9 und 10). Aufgrund der größeren Stichprobenzahl pro Subtest kann die Toleranz hinsichtlich der Abweichung von der mittleren Korrelation nicht mehr so groß sein. Daher kann die Homogenität nur für zwei Konstellationen beibehalten werden (Zeitvorgabe „einzel“ mit $r_+ = .63$ und Antwortformat „weiter“ mit $r_+ = .54$).

10.8 Vergleich der Varianzen

Die Leistung in einem Subtest wird in den meisten Bedingungen durch die Wiederholung der Testung beeinflusst. Im Idealfall läge dieser Effekt zwar nicht vor. Bei Gleichheit der Varianz und Rangreihenfolge der Versuchspersonen (also hohe Korrelationen, vgl. Tabelle 10.29) wird die Verschiebung der Mittelwerte jedoch beherrschbar. In Tabelle 10.31 werden die Ergebnisse der Vergleiche der Varianzen der PBV und der CV dargestellt. Die deskriptiven Werte der Varianzen sind in Anhang 15.7, S. 203 tabelliert, zusammenfassend auch in Tabelle 10.6 und Tabelle 10.18). In 37 der 155 Paarvergleiche bleibt die Prüfgröße T_V (vgl. Formel (7-14)) unter dem jeweiligen kritischen Wert für $\alpha = .10$

Tabelle 10.31: Vergleich der Varianzen von PBV und CV in den 15 Bedingungen durch Morgans T-2V-Test (1939; vgl. Formel (7-14); T_V , in Klammern p -Value und n)

Präsentation		Einzel									Alle					
Antwort-format		Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar	
Zeitvorgabe		Ohne	Gesamt	Einzel	Ohne	Gesamt	Einzel	Ohne	Gesamt	Einzel	Ohne	Gesamt	Ohne	Gesamt	Ohne	Gesamt
Bedingung		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Subtest	1	-37.61 ($<.01$, 18)	5.05 ($<.01$, 15)	7.89 ($<.01$, 17)	-6.83 ($<.01$, 17)	-4.99 ($<.01$, 16)	-6.35 ($<.01$, 17)	9.71 ($<.01$, 19)	13.07 ($<.01$, 15)	-17.09 ($<.01$, 15)	-7.39 ($<.01$, 19)	-10.4 ($<.01$, 15)	-1.20 (.25, 14)	-0.46 (.65, 16)	-15.9 ($<.01$, 17)	-5.60 ($<.01$, 16)
	2	-0.34 (.74, 16)	-12.68 ($<.01$, 18)	15.28 ($<.01$, 15)	-4.01 ($<.01$, 17)	-7.39 ($<.01$, 17)	2.45 (.03, 16)	-11.49 ($<.01$, 17)	-16.11 ($<.01$, 19)	-1.63 (.12, 15)	-9.14 ($<.01$, 15)	-4.28 ($<.01$, 19)	-11.0 ($<.01$, 15)	-2.05 (.06, 13)	-20.3 ($<.01$, 17)	-13.8 ($<.01$, 17)
	3	0.68 (.50, 17)	5.65 ($<.01$, 16)	-11.81 ($<.01$, 18)	1.85 (.08, 15)	0.04 (.97, 17)	-4.09 ($<.01$, 17)	2.85 (.01, 16)	-2.61 (.02, 17)	-4.66 ($<.01$, 19)	4.15 ($<.01$, 15)	4.01 ($<.01$, 14)	-7.41 ($<.01$, 19)	-2.85 (.01, 15)	-1.93 (.08, 13)	-1.68 (.11, 17)
	4	0.28 (.78, 17)	-4.60 ($<.01$, 17)	3.76 ($<.01$, 16)	-9.49 ($<.01$, 18)	-1.75 (.10, 15)	0.95 (.35, 17)	2.09 (.05, 17)	-9.43 ($<.01$, 16)	1.65 (.12, 17)	1.25 (.23, 19)	0.33 (.75, 15)	-4.11 ($<.01$, 15)	-4.83 ($<.01$, 19)	-1.40 (.18, 15)	-0.36 (.72, 14)
	5	1.79 (.10, 14)	4.94 ($<.01$, 17)	3.70 ($<.01$, 17)	2.97 ($<.01$, 16)	-8.78 ($<.01$, 18)	-6.38 ($<.01$, 15)	-3.81 ($<.01$, 16)	-0.34 (.74, 17)	-0.04 (.97, 16)	1.07 (.30, 16)	0.34 (.74, 19)	1.96 (.07, 15)	1.47 (.16, 15)	-5.99 ($<.01$, 19)	-2.99 ($<.01$, 15)
	6	2.07 (.06, 15)	9.10 ($<.01$, 14)	7.59 ($<.01$, 17)	-1.73 (.10, 17)	10.27 ($<.01$, 16)	8.72 ($<.01$, 18)	4.50 ($<.01$, 15)	-4.28 ($<.01$, 16)	9.72 ($<.01$, 17)	-1.50 (.15, 15)	-3.01 ($<.01$, 17)	4.07 ($<.01$, 19)	-2.48 (.03, 15)	3.89 ($<.01$, 15)	-3.77 ($<.01$, 19)
	7	7.42 ($<.01$, 19)	-16.00 ($<.01$, 14)	5.04 ($<.01$, 14)	-8.14 ($<.01$, 17)	-5.49 ($<.01$, 17)	11.76 ($<.01$, 16)	0.81 (.43, 18)	5.37 ($<.01$, 15)	-1.49 (.16, 16)	3.16 ($<.01$, 13)	0.65 (.53, 14)	10.08 ($<.01$, 16)	16.73 ($<.01$, 19)	7.47 ($<.01$, 15)	-2.71 (.02, 15)
	8	4.92 ($<.01$, 15)	0.10 (.92, 19)	-8.12 ($<.01$, 14)	0.52 (.61, 14)	0.60 (.56, 17)	-5.11 ($<.01$, 17)	2.03 (.06, 16)	-3.62 ($<.01$, 18)	-3.74 ($<.01$, 15)	14.31 ($<.01$, 16)	5.13 ($<.01$, 17)	-12.2 ($<.01$, 14)	-14.2 ($<.01$, 17)	-17.5 ($<.01$, 19)	-1.40 (.18, 15)
	9	4.92 ($<.01$, 17)	-13.98 ($<.01$, 15)		-2.13 (.05, 13)	-1.17 (.25, 29)		-0.15 (.88, 30)	-0.80 (.44, 14)				11.70 ($<.01$, 32)	1.32 (.20, 30)	19.03 ($<.01$, 33)	10.09 ($<.01$, 18)
	10	-8.61 ($<.01$, 18)	18.48 ($<.01$, 16)		0.69 (.50, 17)	-23.25 ($<.01$, 28)		-8.12 ($<.01$, 28)	-2.36 (.02, 32)				9.13 ($<.01$, 14)	17.90 ($<.01$, 31)	20.81 ($<.01$, 14)	22.98 ($<.01$, 30)
	11	-8.08 ($<.01$, 15)	4.09 ($<.01$, 16)	-16.77 ($<.01$, 16)	6.49 ($<.01$, 14)	-7.66 ($<.01$, 17)	-1.77 (.09, 18)	-3.06 ($<.01$, 13)	9.35 ($<.01$, 13)	-9.30 ($<.01$, 16)	-4.74 ($<.01$, 13)	-0.03 (.98, 12)	0.76 (.46, 14)	-2.35 (.03, 14)	2.82 (.02, 12)	10.57 ($<.01$, 17)

Bemerkungen: Nicht signifikante Varianzunterschiede mit $p > .10$ grau markiert. Negative T_V resultieren, wenn $S_{CV}^2 > S_{PBV}^2$. p -Value für $|T_V|$ mit $df = n - 2$.

Wird für den Vergleich der Varianzen nicht deren Differenz betrachtet, sondern mittels Wilks L_{vc} (1946; vgl. Formel (7-21)) das Verhältnis, kehrt sich das Verhältnis von als unterschiedlich anzunehmenden Varianz-Vergleichen um. In Tabelle 10.32 werden die logarithmierten Prüfgrößen angegeben, die bei 25 der 155 Vergleiche größer als der kritische Wert sind. Im Unterschied zur Prüfung mit T_V wird hier also bei 130 Vergleichen Varianzgleichheit angenommen. Selbst bei $\alpha = .20$ kann bei 115 Bedingungskonstellationen die Annahme der Gleichheit der Varianzen beibehalten werden.

Tabelle 10.32: Vergleich der Varianzen und Kovarianzen von PBV und CV in den 15 Bedingungen durch Wilks L_{vc} -Test (1946; vgl. Formel (7-21)); $\chi^2(L_{vc}, 1) = -n \ln L_{vc}$, in Klammern p -Value und n)

Präsentation		Einzel									Alle					
Antwort-format		Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar	
Zeitvorgabe		Ohne	Gesamt	Einzel	Ohne	Gesamt	Einzel	Ohne	Gesamt	Einzel	Ohne	Gesamt	Ohne	Gesamt	Ohne	Gesamt
Bedingung		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Subtest	1	3.89 (.05, 18)	0.30 (.58, 15)	0.18 (.67, 17)	0.62 (.43, 17)	0.20 (.66, 16)	0.57 (.45, 17)	0.59 (.44, 19)	0.48 (.49, 15)	3.54 (.06, 15)	0.53 (.47, 19)	0.47 (.49, 15)	0.01 (.94, 14)	0.00 (.98, 16)	2.53 (.11, 17)	0.53 (.47, 16)
	2	0.00 (.96, 16)	1.42 (.23, 18)	1.80 (.18, 15)	0.35 (.56, 17)	0.62 (.43, 17)	0.09 (.76, 16)	0.77 (.38, 17)	2.47 (.12, 19)	0.01 (.91, 15)	0.72 (.40, 15)	0.12 (.73, 19)	2.80 (.09, 15)	0.19 (.66, 13)	3.07 (.08, 17)	1.43 (.23, 17)
	3	0.03 (.86, 17)	3.35 (.07, 16)	6.09 (.01, 18)	0.27 (.60, 15)	0.00 (.99, 17)	0.67 (.41, 17)	0.95 (.33, 16)	0.47 (.49, 17)	1.18 (.28, 19)	1.51 (.22, 15)	1.78 (.18, 14)	2.58 (.11, 19)	1.20 (.27, 15)	0.43 (.51, 13)	0.45 (.50, 17)
	4	0.00 (.96, 17)	1.39 (.24, 17)	2.38 (.12, 16)	13.57 (.01, 18)	0.38 (.54, 15)	0.02 (.89, 17)	1.21 (.27, 17)	3.24 (.07, 16)	0.60 (.44, 17)	0.13 (.71, 19)	0.02 (.89, 15)	1.07 (.30, 15)	1.29 (.26, 19)	0.22 (.64, 15)	0.03 (.86, 14)
	5	0.40 (.53, 14)	2.70 (.10, 17)	1.62 (.20, 17)	0.44 (.51, 16)	5.76 (.02, 18)	1.56 (.21, 15)	0.52 (.47, 16)	0.00 (.95, 17)	0.00 (.99, 16)	0.15 (.70, 16)	0.00 (.95, 19)	0.29 (.59, 15)	0.26 (.61, 15)	4.07 (.04, 19)	0.54 (.46, 15)
	6	0.09 (.77, 15)	0.61 (.44, 14)	0.70 (.40, 17)	0.01 (.93, 17)	1.02 (.31, 16)	0.97 (.32, 18)	0.23 (.63, 15)	0.15 (.70, 16)	1.80 (.18, 17)	0.03 (.85, 15)	0.20 (.65, 17)	0.51 (.48, 19)	0.05 (.83, 15)	0.41 (.52, 15)	0.17 (.68, 19)
	7	2.42 (.12, 19)	3.98 (.05, 14)	0.35 (.55, 14)	4.06 (.04, 17)	1.71 (.19, 17)	4.19 (.04, 16)	0.04 (.85, 18)	0.82 (.36, 15)	0.09 (.76, 16)	0.13 (.72, 13)	0.01 (.94, 14)	2.43 (.12, 16)	2.70 (.10, 19)	4.52 (.03, 15)	0.12 (.72, 15)
	8	0.77 (.38, 15)	0.00 (.99, 19)	1.49 (.22, 14)	0.04 (.85, 14)	0.01 (.93, 17)	0.85 (.36, 17)	0.10 (.75, 16)	0.49 (.48, 18)	0.21 (.65, 15)	2.75 (.10, 16)	1.06 (.30, 17)	2.27 (.13, 14)	3.54 (.06, 17)	6.15 (.01, 19)	0.08 (.78, 15)
	9	1.90 (.17, 17)	9.50 (.01, 15)		0.23 (.63, 13)	0.09 (.76, 29)		0.00 (.97, 30)	0.04 (.84, 14)				7.58 (.01, 32)	0.07 (.79, 30)	12.20 (.01, 33)	6.65 (.01, 18)
	10	1.01 (.31, 18)	9.05 (.01, 16)		0.01 (.92, 17)	12.65 (.01, 28)		1.34 (.25, 28)	0.07 (.79, 32)				1.66 (.20, 14)	3.07 (.08, 31)	5.13 (.02, 14)	11.00 (.01, 30)
	11	1.39 (.24, 15)	0.06 (.81, 16)	0.88 (.35, 16)	0.41 (.52, 14)	1.89 (.17, 17)	0.06 (.81, 18)	0.26 (.61, 13)	0.67 (.41, 13)	1.48 (.22, 16)	0.29 (.59, 13)	0.00 (1.0, 12)	0.01 (.93, 14)	0.08 (.78, 14)	0.02 (.90, 12)	1.07 (.30, 17)

Bemerkungen: Signifikante Varianzunterschiede mit $p < .10$ grau, mit $p < .20$ hellgrau markiert.

10.9 Schlussfolgerungen für die einzelnen Subtests

Sowohl aus den Mittelwertsvergleichen als auch aufgrund der Zusammenhangsanalysen muss gefolgert werden, dass sich eine generelle Bedingungskonstellation als Empfehlung für die Kon-

struktion einer Computerversion nicht ableiten lässt. Insgesamt ist vielmehr die Schlussfolgerung zu ziehen, dass ein leistungsverringender Effekt durch die Durchführung eines Leistungstests am PC entsteht. Trotz der geringen Fallzahlen in den Bedingungen pro Subtest sollen jedoch abschließend für die einzelnen Subtests basierend auf ihren inhaltlichen Merkmalen Empfehlungen für weitere Entwicklungen abgeleitet werden. Zur Übersicht wird in Tabelle 10.33 die gleichzeitige Überprüfung von Mittelwert, Varianz und Kovarianz nach Wilks (1946; vgl. Formel (7-19) (ohne Berücksichtigung der Testreihenfolge) angeführt. Wie sich bereits in Kapitel 10.8 zeigte, ist dieser Test eher liberal als konservativ, weshalb $\alpha = .20$ gewählt wurde.

Tabelle 10.33: Vergleich der Mittelwerte, Varianzen und Kovarianzen von PBV und CV in den 15 Bedingungen durch Wilks L_{mvc} -Test (1946; vgl. Formel (7-19)); $\chi^2(L_{mvc}, 1) = -n \ln L_{mvc}$, in Klammern p -Value und n)

Präsentation		Einzel									Alle					
Antwort-format		Einmalig			Weiter			Korrigierbar			Einmalig		Weiter		Korrigierbar	
Zeitvorgabe		Ohne	Gesamt	Einzel	Ohne	Gesamt	Einzel	Ohne	Gesamt	Einzel	Ohne	Gesamt	Ohne	Gesamt	Ohne	Gesamt
Bedingung		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Subtest	1	6.86 (.01, 18)	3.54 (.06, 15)	3.02 (.08, 17)	5.20 (.02, 17)	3.64 (.06, 16)	4.72 (.03, 17)	10.32 (.01, 19)	3.90 (.05, 15)	7.38 (.01, 15)	7.53 (.01, 19)	3.76 (.05, 15)	2.76 (.10, 14)	3.67 (.06, 16)	8.76 (.01, 17)	2.71 (.10, 16)
	2	2.26 (.13, 16)	5.55 (.02, 18)	4.46 (.03, 15)	5.76 (.02, 17)	1.98 (.16, 17)	4.13 (.04, 16)	2.54 (.11, 17)	3.45 (.06, 19)	2.30 (.13, 15)	3.71 (.05, 15)	6.17 (.01, 19)	7.13 (.01, 15)	2.88 (.09, 13)	5.67 (.02, 17)	3.14 (.08, 17)
	3	3.39 (.07, 17)	1.40 (.24, 16)	1.95 (.16, 18)	3.33 (.07, 15)	4.39 (.04, 17)	2.55 (.11, 17)	3.58 (.06, 16)	2.37 (.12, 17)	6.96 (.01, 19)	7.44 (.01, 15)	2.75 (.10, 14)	6.68 (.01, 19)	2.43 (.12, 15)	2.23 (.13, 13)	3.52 (.06, 17)
	4	0.21 (.65, 17)	2.51 (.11, 17)	1.66 (.20, 16)	4.90 (.03, 18)	2.51 (.11, 15)	0.25 (.62, 17)	0.52 (.47, 17)	2.70 (.10, 16)	2.41 (.12, 17)	0.17 (.68, 19)	0.78 (.38, 15)	0.32 (.57, 15)	0.59 (.44, 19)	1.38 (.24, 15)	1.12 (.29, 14)
	5	1.18 (.28, 14)	1.05 (.30, 17)	1.98 (.16, 17)	3.71 (.05, 16)	2.36 (.12, 18)	0.94 (.33, 15)	3.02 (.08, 16)	0.39 (.53, 17)	3.83 (.05, 16)	0.83 (.36, 16)	1.39 (.24, 19)	3.26 (.07, 15)	3.64 (.06, 15)	4.25 (.04, 19)	0.21 (.65, 15)
	6	2.49 (.11, 15)	2.96 (.09, 14)	4.81 (.03, 17)	3.12 (.08, 16)	6.52 (.01, 17)	3.31 (.07, 18)	5.34 (.02, 15)	6.16 (.01, 16)	5.68 (.02, 17)	4.83 (.03, 15)	5.01 (.03, 17)	4.64 (.03, 19)	1.67 (.20, 15)	2.50 (.11, 15)	3.07 (.08, 19)
	7	2.66 (.10, 19)	7.36 (.01, 14)	2.43 (.12, 14)	6.30 (.01, 17)	2.14 (.14, 17)	1.78 (.18, 16)	5.93 (.01, 18)	6.40 (.01, 15)	4.62 (.03, 16)	0.96 (.33, 13)	4.30 (.04, 14)	3.14 (.08, 16)	8.78 (.01, 19)	6.34 (.01, 15)	2.45 (.12, 15)
	8	2.69 (.10, 15)	4.21 (.04, 19)	2.32 (.13, 14)	3.31 (.07, 17)	3.61 (.06, 17)	6.94 (.01, 17)	3.13 (.08, 16)	4.22 (.04, 18)	5.31 (.02, 15)	2.06 (.15, 16)	3.97 (.05, 17)	2.72 (.10, 14)	4.25 (.04, 17)	10.25 (.01, 19)	3.90 (.05, 15)
	9	2.71 (.10, 17)	3.67 (.06, 15)		1.47 (.23, 13)	2.19 (.14, 29)		11.86 (.01, 30)	4.47 (.03, 14)				7.15 (.01, 32)	0.31 (.57, 30)	11.43 (.01, 33)	4.58 (.03, 18)
	10	1.60 (.21, 18)	15.23 (.01, 16)		4.98 (.03, 17)	9.35 (.01, 28)		14.83 (.01, 28)	5.66 (.02, 32)				16.16 (.01, 14)	38.34 (.01, 31)	19.09 (.01, 14)	33.10 (.01, 30)
	11	3.67 (.06, 15)	2.78 (.10, 16)	1.02 (.31, 16)	1.69 (.19, 14)	4.54 (.03, 17)	0.10 (.76, 18)	3.13 (.08, 13)	1.95 (.16, 13)	3.12 (.08, 16)	2.87 (.09, 13)	2.26 (.13, 12)	0.49 (.48, 14)	0.68 (.41, 14)	7.99 (.01, 12)	1.50 (.22, 17)

Bemerkungen: Nicht signifikante Unterschiede mit $p < .20$ (wegen des liberalen Tests) sind grau markiert.

Subtest 1: Bei Mittelung der beiden Durchführungsbedingungen (und somit der Reihenfolgeeffekte) weist der geringste Effekte den Wert -0.18 auf, wenn die Items einzeln präsentiert werden, nach Abgabe der Antwort ein „weiter“-Button geklickt werden muss und die Gesamtbearbeitungszeit angezeigt wird (Bedingung 5, Tabelle 10.24). Die Test-Retest-Korrelation ist in diesem Fall .66 ($n = 16$, Tabelle 10.29). Der Vergleich der PBV zu T1 mit der CV zu T1 resultiert in

Bedingung 5 jedoch in einem Effekt von $d_{1,5} = -4.93$ (Tabelle 10.27). Für diesen Vergleich zeigen die Bedingungen 12 und 13 die geringsten Werte (alle Items gleichzeitig, Antworten weiterhin sicht-, aber nicht mehr korrigierbar, mit und ohne Anzeige der Gesamtbearbeitungsdauer), allerdings mit entgegengesetzten Vorzeichen ($d_{1,12} = -0.52$; $d_{1,13} = 0.44$). Nach dem L_{mvc} -Test muss für jede Bedingung echte Äquivalenz abgelehnt werden (Tabelle 10.33). Hinsichtlich Korrelation und Varianzgleichheit zwischen PBV und CV zeigt Bedingung 13 jedoch gute Werte (Tabelle 10.29 und Tabelle 10.31). Insgesamt zeigt sich, dass die Anzeige der Gesamtbearbeitungsdauer hinsichtlich der Äquivalenz sowohl korrelativ als auch bzgl. des Mittelwertvergleichs vorteilhaft für Subtest 1 ist.

Subtest 2: Im Gesamtvergleich liegen fünf Effektgrößen kleiner .15 vor, beim Vergleich der Modi zu T1 ist die Differenz der Mittelwerte lediglich in Bedingung 5 unter dieser Grenze ($d_{2,5} = -0.03$, Tabelle 10.27); PBV und CV korrelieren dann mit .59 ($n = 17$). Von den drei Bedingungen mit Korrelationen größer .70 (7, 9 und 11) zeigt nur Bedingung 9 noch einigermaßen kleine Effektgrößen ($d_{PBV,CV} = -0.20$, $d_{T1} = -0.36$). Insgesamt erweist sich Subtest 2 als weniger anfällig hinsichtlich Verzerrungen durch den Bearbeitungsmodus, wenngleich eine eindeutig äquivalente Bedingung auch hier nicht beobachtbar ist (Tabelle 10.33). Für die beiden sprachlichen Subtests zeigen sich jedoch Tendenzen, dass die bisherige Methodik der Darstellung der Items am PC zu den geringsten Verzerrungen führt.

Subtest 3: Die aufgrund der kognitiven Anforderung erwartbare Ähnlichkeit der psychometrischen Merkmale der Subtests 3, 4 und 5 spiegelt sich in den Ergebnissen nicht wieder. In Bedingung 4 weist Subtest 3 sowohl im Gesamtvergleich, als auch in den Ergebnissen zu T1 kleine Werte auf ($d_{PBV,CV} = -0.14$, $d_{T1} = 0.03$), allerdings ist auch die Test-Retest-Korrelation klein. In Bedingung 12 ist die Korrelation von PBV und CV mit .72 ebenso wie der Gesamteffekt $d_{PBV,CV} = -0.14$ akzeptabel; zu T1 sind die Werte in der PBV allerdings deutlich höher als in der CV ($d_{T1} = -0.76$). Der Overall-Test (Tabelle 10.33) würde auf Äquivalenz in Bedingung 2 hindeuten. Hierbei ist jedoch die Test-Retest-Korrelation deskriptiv negativ. Es zeigt sich erneut, dass der L_{mvc} -Test ein eher liberales Kriterium darstellt, insbesondere beim Vergleich von zwei Tests, da dann die Korrelation nur sehr geringen Einfluss auf die Höhe der Prüfgröße hat.

Subtest 4: Der L_{mvc} -Test wird für neun der 15 Bedingungen nicht signifikant (Tabelle 10.33). Bedingung 1 (Items werden einzeln präsentiert, bei Beantwortung erscheint das nächste Item, keine Zeitanzeige) weist sowohl in der Gesamtbetrachtung ($d_{PBV,CV} = 0.05$), als auch im Vergleich der Testungen zu T1 ($d_{T1} = 0.04$) sowie in der Korrelation von PBV und CV ($r = .80$, $n = 17$) Werte auf, die auf Äquivalenz hindeuten. Die Varianzen sind in beiden Fällen statistisch nicht verschieden: $S_{CV \text{ zu } T1}^2 = 13.03$, $S_{PBV \text{ zu } T1}^2 = 5.90$, $F(10,5) = 2.21, p.20$; $S_{CV}^2 = 11.6$, $S_{PBV}^2 = 11.42$, $|T_V(15)| = 0.28, p = .78$ (vgl. Tabelle 10.31). Korrelativ zeigen auch die Bedingungen 6, 8 und 13 für Subtest 4 Werte größer .70 (Tabelle 10.29). In diesen Bedingungen zeigen sich jedoch in den jeweiligen Mittelwertvergleichen vereinzelt stärkere Effekte. Es bleibt festzuhalten, dass die Äquivalenz der Subtests in diesen Bedingungskonstellationen prinzipiell funktionieren könnte.

Subtest 5: Auch hier werden sieben der 15 Bedingungen im L_{mvc} -Test statistisch nicht signifikant (Tabelle 10.33). Die am ehesten kongruente Bedingung hinsichtlich der drei Kriterien ist Bedingung 9. Im Vergleich der Testung zu T1 zeigen die Modi keinen Unterschied. Über beide Testungen hinweg ist der Unterschied jedoch zu hoch ($d_{PBV,CV} = -0.25$) und die Korrelation zu

niedrig ($r = .54, n = 19$). Test-Retest-Korrelationen größer .70 liegen in den Bedingungen 8 und 11 vor; ein geringer Mittelwertsunterschied in Bedingung 2. Insgesamt lässt sich kein Trend identifizieren, der eine der Bedingungen favorisieren würde.

Subtest 6: Die Ergebnisse des Overall-Tests werden in allen Bedingungen statistisch signifikant. Bedingung 1 zeigt für den Gesamtvergleich quasi keinen Unterschied zwischen PBV und CV ($d_{PBV,CV} = -0.01$) und für den Vergleich zu T1 zwar den geringsten, aber dennoch einen deutlichen Unterschied ($d_{T1} = -0.17$); die Test-Retest-Korrelation ist aber zu niedrig und statistisch nicht von Null verschieden ($r = .43, n = 15$). In sechs Bedingungen liegen Korrelationen größer .70 vor (2, 3, 4, 8, 13 und 15). Zu T1 ist der Unterschied in Bedingung 3 noch relativ gering ($d_{T1} = -0.20$), im Gesamtvergleich betragen die kleinsten Effekte -0.24 (Bedingung 4 und 15). Hinsichtlich der Reihenfolge in PBV und CV scheint die Gesamtzeitanzeige förderlich zu sein. Bezüglich gleicher Leistungen lässt nur Bedingung 1 positivere Erwartungen zu.

Subtest 7: Der L_{mvc} -Test wird für Bedingung 10 nicht signifikant, obwohl die Varianzen sich hier dem T_V -Test nach unterscheiden. Es lassen sich jedoch zwei weitere Bedingungen ausmachen, deren Ergebnisse auf Äquivalenz hindeuten (6 und 11), die hinsichtlich der Faktorausprägungen jedoch nicht unterschiedlicher sein könnten: $d_{PBV,CV}^6 = 0.11$, $d_{T1}^6 = -0.05$, $r^6 = .69$, $n^6 = 16$; $d_{PBV,CV}^{11} = -0.06$, $d_{T1}^{11} = -0.10$, $r^{11} = .77$, $n^{11} = 14$. Für Bedingung 6 ist die Gleichheit der Varianzen nicht anzunehmen: $S_{CV \text{ zu } T1}^2 = 13.61$, $S_{PBV \text{ zu } T1}^2 = 1.87$, $F(9,5) = 13.61, p = .02$; $S_{CV}^2 = 9.98$, $S_{PBV}^2 = 23.72$, $|T_V(14)| = 11.76, p = < .01$. In Bedingung 11 sind die Varianzen zwar höher, jedoch wesentlich ähnlicher: $S_{CV \text{ zu } T1}^2 = 28.99$, $S_{PBV \text{ zu } T1}^2 = 32.03$, $F(11,14) = 1.04, p = .42$; $S_{CV}^2 = 25.45$, $S_{PBV}^2 = 26.36$, $|T_V(12)| = 0.65, p = .53$, $-14 \ln L_{vc} = 0.01, p = .94$.

Subtest 8: Die beiden Bedingungen bei der Gesamtbetrachtung, deren Effekte unter dem Kriterium von ± 0.15 liegen (1 und 2), zeigen eine mäßig hohe Test-Retest-Korrelation. Bedingung 7 (einzeln, korrigierbar, ohne Zeitanzeige) weist zwar für die Gesamtbetrachtung einen etwas erhöhten Effekt von -0.22 auf; im Vergleich der Modi zu T1 besteht jedoch fast kein Unterschied (-0.01); zudem ist die Korrelation der Modi akzeptabel ($r = .71, n = 16$). Allerdings unterscheiden sich die Varianzen zu T1 deutlich ($S_{CV \text{ zu } T1}^2 = 11.29$, $S_{PBV \text{ zu } T1}^2 = 38.97$, $F(9,5) = 3.45, p = .09$, $S_{CV}^2 = 38.74$, $S_{PBV}^2 = 20.52$, $|T_V(14)| = 2.03, p = .06$).

Subtest 9: Dies ist der einzige Subtest, bei dem in der Gesamtbetrachtung und im Vergleich der Testungen zu T1 in der Mehrheit der Bedingungen höhere Werte in der CV erbracht wurden; außer in den Bedingungen, in denen die Items einzeln präsentiert wurden und ein „weiter“-Button geklickt werden musste (4, 5, 7 und 8). Offenbar spielen die Zeitverluste hier eine entscheidende Rolle. Insbesondere die Test-Retest-Korrelationen fallen in allen Bedingungen sehr niedrig aus; die höchsten Werte sind in den Bedingungen 12, 13 und 14, mit $r^{14} = .59$ ($n^{14} = 33$) als höchsten Wert. Die Modi unterscheiden sich zu T1 in allen Bedingungen. Nur Bedingung 2 in der Gesamtbetrachtung weist mit 0.05 einen annähernd niedrigen Wert auf. Hinsichtlich der Varianz zeigt insbesondere Bedingung 13 akzeptable Werte ($S_{CV}^2 = 18.13$, $S_{PBV}^2 = 19.63$, $|T_V(28)| = 1.32, p = .20$, $-30 \ln L_{vc} = 0.07, p = .79$); trotz einer Effektstärke von $d_{PBV,CV}^{13} = .76$ deutet sogar der L_{mvc} -Test auf Äquivalenz hin ($-30 \ln L_{vc} = 0.31, p = .57$).

Subtest 10: Ähnlich wenige Hinweise auf Äquivalenz förderliche Bedingungen liegen für Subtest 10 vor. Obwohl der Overall-Test für Bedingung 1 Äquivalenz anzeigt ($-18 \ln L_{vc} = 1.60, p = .21$), sprechen die einzelnen Parameter eher dagegen ($d_{PBV,CV}^1 = -.52$, $|T_V(16)| = 8.61, p = <$

.01, $r^1 = .38$); nur beim Vergleich der Leistungen zu T1 liegt eine kleine Effektgröße vor ($d_{T1} = 0.03$). Die Gesamtdarstellung am PC führt zu extremen Unterschieden im Vergleich zur PBV. Dies könnte durch die größere Darstellung der Items am PC und die damit einhergehende längere Bearbeitungsdauer über alle Items hinweg bedingt sein. In Bedingung 13 und 14 entsprechen sich immerhin die Probandenreihenfolgen in angemessener Höhe ($r^{13} = .75$, $n^{14} = 31$; $r^{14} = .71$, $n^{14} = 14$), die Varianzen unterscheiden sich hingegen deutlich (vgl. Tabelle 10.31). Insgesamt bleibt festzuhalten, dass der Speedcharakter der Subtests 9 und 10 wohl das größte Problem hinsichtlich der Äquivalenz der beiden Modi darstellt (anders als z. B. in den Untersuchungen von Neuman und Baydoun (1998) mit dem OST; vgl. Kapitel 4.1).

Subtest 11: Hinsichtlich des Mittelwertes liegen in der Gesamtbetrachtung und im Vergleich der Testungen zu T1 für die Bedingungen 3 und 8 geringe Effekte 0.15 vor. Auch die Test-Rest-Korrelationen sind in diesen Bedingungen größer .80. Jedoch unterscheiden sich die Varianzen dem $T - 2V$ -Test nach deutlich ($|T_V^3(14)| = 16.77, p = < .01$, $|T_V^8(11)| = 9.35, p = < .01$). Der L_{mvc} -Test würde hingegen für Äquivalenz sprechen (Tabelle 10.33). Varianzgleichheit liegt für die Bedingungen 11 und 12 vor, wobei die Test-Rest-Korrelation nur für Bedingung 11 größer .70 ist. Insgesamt zeigen sich durchwegs einigermaßen hohe Ausprägungen hinsichtlich der Korrelationen und geringe Effekte bei den Mittelwerten, so dass mit Subtest 11 mit Ausnahme der Varianzen die Äquivalenzkriterien in den meisten Bedingungen eingehalten werden könnten.

11. Die Struktur des neuen Leistungsprüfsystems

Die in Kapitel 3 vorgestellte Einbettung der Struktur des LPS-neu in das Intelligenzmodell von Carroll (1993) wird im Folgenden auf seine statistische Passung hin überprüft. Während Carrolls Ergebnisse auf den Analysen mittels der explorativen hierarchischen Faktorenanalyse nach Schmid und Leimann (1957) basieren, wird die Struktur des LPS-neu mit Hilfe einer konfirmatorischen Faktorenanalyse (CFA) überprüft. Als Vorlage dient ein „Second-Order CFA Model“, das von Byrne (2001) in einem Lehrbuch zur Nutzung von AMOS (Analysis of Moment Structures; Arbuckle, 2008a) vorgestellt wird. Im allgemeinen Vorgehen wird dem Lehrbuch von Kline (2004) und den Vorgaben für die Darstellung von Strukturgleichungsmodellen (SEM) von Boomsma (2000) und McDonald und Ho (2002) gefolgt.

11.1 Datenaufbereitung

Nach Boomsma (2000) stehen zu Beginn die Darstellung des Modells und die theoretische Fundierung der latenten Variablen. Dies wurde bereits in Kapitel 3 abgehandelt. Nach der Datenerhebung sind die Aufbereitung der Daten und ein Screening erforderlich (Kline, 2004). Da die Rohdaten zur Verfügung stehen, werden diese dem Programm eingegeben. Weil die folgende Analyse auf einem vollständigen Datensatz beruhen soll, wird als nächstes nach Ausreißern gefahndet. Kline (2004) empfiehlt hierfür, die Mahalanobis Distanz, also den Abstand eines n -dimensionalen Punktes vom n -dimensionalen Mittelwert (Centroid):

$$(11-1) \quad d^2 = (\mathbf{X}^n - \boldsymbol{\mu}^n)' \boldsymbol{\Sigma}^{-1} (\mathbf{X}^n - \boldsymbol{\mu}^n)$$

Im vorliegenden Fall wird somit untersucht, ob das Leistungsprofil eines Teilnehmers auffällig weit vom durchschnittlichen Profil entfernt ist. Der größte Abstand entspricht einem Wert von 41.98 (bei einem durchschnittlichen Abstand von 17.22, SD = 5.43). Die Rohwerte des Probanden (Subtest 1-11: 18, 4, 24, 8, 14, 11, 20, 4, 1, 11, 3) weisen ihn als insgesamt schwach aus, das Profil erscheint aber weder absichtlich noch durch generelle Unmotiviertheit verfälscht; es zeigt vielmehr ein Defizit in der visuellen Diskriminationsfähigkeit aufgrund der geringen Leistungen in den Subtests 2 und 8 sowie eine mangelhafte Konzentrationsfähigkeit (Subtests 9 und 11). Ebenso ergab die Inspektion der weiteren Probanden mit hohen Distanzen keine zwingenden Gründe, einen der $n = 273$ Probanden, die die Papier-Bleistift-Version (PBV) des LPS-neu vollständig bearbeitet haben, auszuschließen.

Bei der Komprimierung der Informationsdarstellung muss zwischen Korrelations- und Kovarianzmatrix entschieden werden, wobei im Falle der Korrelationsmatrix durch Angabe der Standardabweichungen beide Matrizen ineinander überführbar sind. Zudem sollten die Mittelwerte der manifesten Variablen (hier also der Subtestrohwerte) angegeben werden, um die Möglichkeit zu bieten, Modelle mit Mittelwertsvergleichen rechnen zu können. Um allen Anforderungen gerecht zu werden, werden in Tabelle 11.1 alle Angaben zusammengefasst.

Tabelle 11.1: Interkorrelationen, Kovarianzen, Mittelwerte, Standardabweichungen, Schiefe- und Kurtosisparameter der elf Subtests der PBV des LPS-neu

	1	2	3	4	5	6	7	8	9	10	11
Subtest 1	96.13	35.67	7.89	5.96	11.98	22.12	10.73	13.68	8.36	23.95	23.48
Subtest 2	.48	57.36	9.96	8.67	11.95	21.91	12.65	13.13	8.77	22.45	22.16
Subtest 3	.22	.35	13.88	5.09	6.07	7.35	7.43	8.11	4.07	8.77	6.90
Subtest 4	.21	.40	.48	8.06	5.10	6.99	4.66	5.64	3.73	6.43	7.25
Subtest 5	.37	.48	.50	.55	10.76	9.05	7.03	6.69	5.24	7.11	9.88
Subtest 6	.29	.38	.26	.32	.36	59.27	13.20	11.11	5.83	18.41	15.59
Subtest 7	.21	.32	.39	.32	.42	.33	26.64	12.27	5.49	8.24	10.20
Subtest 8	.25	.31	.39	.36	.37	.26	.43	30.93	5.18	11.40	8.02
Subtest 9	.20	.27	.25	.30	.37	.17	.25	.21	18.79	6.70	10.09
Subtest 10	.34	.41	.33	.31	.30	.33	.22	.28	.21	52.12	19.39
Subtest 11	.34	.42	.27	.37	.43	.29	.28	.21	.33	.39	48.45
Mittelwert	28.33	22.78	27.01	16.68	19.06	22.43	30.57	34.43	11.23	33.15	15.99
SD	9.80	7.57	3.73	2.84	3.28	7.70	5.16	5.56	4.33	7.22	6.96
Schiefe	0.34	-0.05	-0.09	-0.2	0.34	0.13	-0.95	-1.55	0.07	-0.03	0.51
Kurtosis	-0.25	-0.66	1.42	0.59	0.17	-0.37	1.4	3.64	0.00	2.51	1.38

Bemerkungen: Die grau markierte Diagonale zeigt die Varianzen an; über der Diagonalen stehen Kovarianzen; unter der Diagonalen Korrelationen; für $n = 273$ sind Korrelationen mit $r > .16$ signifikant auf einem α -Niveau von 1%

McDonald und Ho (2002) fordern zudem die standardisierte Diskrepanz von vorliegenden und vorhergesagten Korrelationen. Da diese abhängig von der noch zu wählenden Analyseverfahren sind, werden sie erst nach deren Auswahl aufgeführt (vgl. Tabelle 11.4).

Als nächstes sollen die Verteilungen der manifesten Variablen, also der Subtestrohre, untersucht werden (Kline, 2004). Dies wurde bereits in Kapitel 0 ausführlich behandelt; der Übersichtlichkeit halber werden Schiefe und Kurtosis in Tabelle 11.1 aufgeführt. Curran, West und Finch (1996) untersuchen die Robustheit verschiedener SEM-Methoden im Kontext von konfirmatorischen Faktorenanalysen. Als moderat nicht normalverteilt bezeichnen sie eine Verteilung mit einer Schiefe von 2.0 und einer Kurtosis von 7.0; die unnormalste Verteilung im vorliegenden Datensatz liegt für Subtest 8 vor, der allerdings für beide Werte unter diesen Grenzen liegt. Zudem haben Curran et al. (1996) gezeigt, dass bei Vorliegen einer solchen moderaten Verletzung der Normalitätsverteilung für $n = 200$ nicht zutreffende Modelle dennoch korrekt als schlecht passend klassifiziert werden.

Da die höchste Korrelation zwischen Subtest 4 und 5 mit .55 deutlich niedriger als der von Kline (2004) formulierte kritische Wert .85 ist, war nicht zu erwarten, dass Multikollinearität bezüglich der elf Subtests des LPS-neu besteht. Die Überprüfung durch Bestimmung der quadrierten multiplen Korrelation zwischen jeder der elf Variablen und den zehn anderen zeigte, dass mit $R_{smc}^2 = .50$ der höchste Wert für Subtest 5 vorlag, der somit deutlich unter der kritischen Grenze von .90 lag. Zudem war in keinem Regressionsmodell die Toleranz ($1 - R_{smc}^2$) kleiner .50 (kritisch, falls kleiner .10) und der Variance Inflation Factor (VIF, $(1 - R_{smc}^2)^{-1}$) für alle Subtests immer zwischen eins und zwei (kritisch, falls größer 10).

Ein weiterer zu überprüfender Punkt ist die Reliabilität der manifesten Variablen. Sowohl interne Konsistenz als auch Split-Half-Reliabilität wurden bereits in Tabelle 10.8 (S. 93) angegeben. Mit Ausnahme von Subtest 4 ($\alpha = .65$, $r_{XX'} = .58$) liegen alle Werte im sehr guten oder wenigstens adäquaten Bereich.

11.2 Modellschätzung

Der nächste Schritt ist die mathematische Bestimmung des Modells. Prinzipiell kann dies sowohl auf graphische als auch auf formale Art in Form von Matrizen geschehen. Da die Struktur vorliegender CFA nicht sonderlich komplex ist, werden die vermuteten Zusammenhänge in Abbildung 3.6 als hinreichend genau dargestellt erachtet. Die darüber hinaus zu formulierenden Fehler- und Residualvariablen sind in der Darstellung des Ergebnisses der Modellprüfung (Abbildung 3.6) ersichtlich. Um die quasi so formulierten linearen Gleichungen lösen zu können, müssen mindestens so viele Parameter aus den Daten bestimmt werden können, wie Parameter zu schätzen sind. Die Datengrundlage ist die Kovarianzmatrix, die bei elf Subtests $11(11 + 1)/2 = 66$ Werte liefert. Geschätzt werden im vorliegenden Modell 26 Parameter. Somit beträgt die Zahl der Freiheitsgrade (df) 40. Das Modell ist somit identifizierbar. Auf die Zahl der zu schätzenden Parameter wird im Folgenden etwas detaillierter eingegangen, da sie nicht ohne weiteres ersichtlich ist.

In vorliegender CFA werden formal die LPS-neu Leistungen durch die latenten Fähigkeiten des Stratum II (Carroll, 1993) bestimmt, die wiederum vom g -Faktor abhängen. Der lineare Zusammenhang im sogenannten Messmodell wird also durch

$$(11-2) \quad X_{LPS1} = b_{LPS1} \times g_c + e_{LPS1}$$

bestimmt. Analog können die Gleichungen der übrigen zehn Subtests mit den entsprechenden Stratum II Faktoren konstruiert werden (vgl. Abbildung 3.6); ebenso wird der Zusammenhang von Stratum II zu Stratum III z. B. für g_c durch die Gleichung

$$(11-3) \quad g_c = b_{g_c} \times g + res_c$$

formuliert. Werden nun die Schätzungen der Form (11-3) in die Gleichungen der Form (11-2) eingefügt, können die daraus resultierenden elf Funktionen in Matrizenform zu einer Schätzgleichung zusammengefasst werden:

$$(11-4) \quad X_{LPS} = \mathbf{B}_{LPS}(\mathbf{b}_{g_{StratumII}}g + \mathbf{res}_{g_{StratumII}}) + \mathbf{e}_{LPS}$$

$$\begin{pmatrix} X_{LPS1} \\ X_{LPS2} \\ X_{LPS3} \\ X_{LPS4} \\ X_{LPS5} \\ X_{LPS6} \\ X_{LPS7} \\ X_{LPS8} \\ X_{LPS9} \\ X_{LPS10} \\ X_{LPS11} \end{pmatrix} = \begin{pmatrix} b_{LPS1} \\ b_{LPS2} \\ & b_{LPS3} \\ & b_{LPS4} \\ & b_{LPS5} \\ & & b_{LPS6} \\ & & b_{LPS7} \\ & & b_{LPS8} \\ & & & b_{LPS9} \\ & & & b_{LPS10} \\ & & & & b_{LPS11} \end{pmatrix} \left(\begin{pmatrix} b_{g_c} \\ b_{g_f} \\ b_{g_v} \\ b_{g_s} \\ b_{g_t} \end{pmatrix} g + \begin{pmatrix} res_{g_c} \\ res_{g_f} \\ res_{g_v} \\ res_{g_s} \\ res_{g_t} \end{pmatrix} \right) + \begin{pmatrix} e_{LPS1} \\ e_{LPS2} \\ e_{LPS3} \\ e_{LPS4} \\ e_{LPS5} \\ e_{LPS6} \\ e_{LPS7} \\ e_{LPS8} \\ e_{LPS9} \\ e_{LPS10} \\ e_{LPS11} \end{pmatrix}$$

Aus (11-4) wird deutlich, dass sowohl die Fehlervarianzen \mathbf{e}_{LPS} der Messungen als auch die Residualvarianzen, also die nicht durch g erklärbaren Anteile der Stratum II Faktoren lediglich additiv wirken. Ihr Regressionsparameter, also der Pfad in der graphischen Darstellung, beträgt somit 1. Damit sind zunächst 16 Restvarianzen als unbekannte Parameter zu schätzen.

Da für die latenten Variablen des Stratum II und Stratum III keine Skalierung vorliegt, wird die Faktorladung auf jeweils eine manifeste Variable auf 1 festgesetzt, so dass der Zusammen-

hang zwischen den manifesten und der latenten Variable jeweils in diesem Verhältnis gesehen wird. Üblicherweise wird der Pfad der ersten Variable auf 1 gesetzt, prinzipiell ist die Auswahl jedoch willkürlich. Auf der ersten Ebene werden somit $b_{LPS1} = b_{LPS3} = b_{LPS6} = b_{LPS9} = b_{LPS11} = 1$ gesetzt. Auf der zweiten Ebene wird der Pfad von g zu g_f auf 1 gesetzt, da z. B. Gustafsson (1984) zeigte, dass g_f , also die fluide Intelligenz, der beste Prädiktor für g ist. Es werden somit neun Regressionsparameter geschätzt. Als 26. Parameter muss die Varianz von g geschätzt werden. Kline (2004) führt Alternativen für die Festsetzung der Parameter (constraint) bzw. um die latenten Variablen zu skalieren auf; im vorliegenden Fall wurden jedoch die üblichsten Annahmen umgesetzt.

Vor der Wahl der Schätzmethode muss noch angemerkt werden, dass mit $g_t \rightarrow LPS11$ der Empfehlung, in hierarchischen Modellen mindestens zwei Indikatoren pro Faktor erster Ordnung (hier: Stratum II) zu formulieren (Kline, 2004), nicht gefolgt werden kann. Durch die Konstantsetzung der Residualvarianz (res_{g_t}) auf 0 und des Regressionsparameters (b_{LPS11}) auf 1 wird g_t jedoch quasi mit X_{LPS11} gleichgesetzt und nur der Systematik wegen angegeben. Formal wird der Einfluss von g auf X_{LPS11} direkt modelliert. McDonald und Ho (2002) beispielsweise empfehlen in diesem Fall die Aufteilung der Variable in zwei Hälften und somit die Konstruktion zweier Indikatoren. Neben dem Umstand, dass dieses Vorgehen im Ergebnis keine relevanten Veränderungen hervorbringt, wird die gewählte Form aufgrund der höheren inhaltlichen Plausibilität beibehalten.

Als Schätzmodell wird das Maximum Likelihood Verfahren (ML) gewählt, da es sich als das robusteste Verfahren herausgestellt hat, insbesondere wenn die Normalverteilungsannahme als nur schwach verletzt angesehen werden kann, und allgemein als die Methode der Wahl angesehen wird (Blunch, 2008). Auch in der Monte Carlo Studie von Curran et al. (1996) resultierten mit der ML für die vorliegende Stichprobengröße ($n = 273$) die adäquatesten Ergebnisse. Hierbei werden nach einer zufälligen Bestimmung der zu schätzenden Parameter diese solange verändert und angepasst, bis die Differenzfunktion zwischen der Kovarianzmatrix, die aus den Daten geschätzt wurde, und der Kovarianzmatrix, die aufgrund der Modellparameter bestimmt wird, minimal ist und somit die „Wahrscheinlichkeit der Daten“ (Maximum Likelihood) maximal. AMOS stützt sich dabei auf die Algorithmen von Browne (1984; Arbuckle, 2008b).

11.3 Modellgüte

Es gibt eine Vielzahl von Indizes, die die Modellgüte der Schätzung angeben, wobei es keinen „magic index“ gibt, der als „gold standard“ (Kline, 2004, S. 134) dienen kann. Daher meinen McDonald und Ho (2002, S. 72): „It is sometimes suggested that we should report a large number of these indices, apparently because we do not know how to use any of them.“ Jeder Index ist für verschiedene Verletzungen der für ihn getroffenen Annahmen unterschiedlich sensibel bzw. dagegen robust. Fittet ein Modell in allen Indizes, ist die Chance, das Modell aus diesem Grund irrtümlich nicht abzulehnen, gering. Daher werden die meisten von AMOS berechneten Indizes angegeben. Kline (2004) warnt aber davor, sich lediglich auf die Fit-Indizes zu verlassen; Boomsma (2000) und McDonald und Ho (2002) betonen, die Interpretation der geschätzten Parameter nicht zu vernachlässigen.

Für das formulierte Modell resultiert ein χ^2_M von 62.48 mit $df = 40$ und $p = .01$. χ^2_M wird in allen Abhandlungen über SEM kritisch diskutiert, da es durch viele Faktoren (z. B. Stichprobengröße, Höhe der Interkorrelationen der manifesten Variablen) beeinflusst werden kann und nicht als besonders aussagekräftig gesehen wird, da sein Wert schnell über die kritische Grenze steigt und somit die Nullhypothese, dass die geschätzte Kovarianzmatrix der empirischen entspricht bzw. das Modell passt, abgelehnt werden muss. Da mit jedem Freiheitsgrad χ^2_M steigt, wurde ein normiertes χ^2 (NC) vorgeschlagen (χ^2_M/df), für das Bollen (1989) Höchstgrenzen von 2.0, 3.0 oder auch 5.0 angeben hat (Kline, 2004). Hierfür übersteigt das Modell mit $NC = 1.56$ den kleinsten kritischen Wert nicht. Da aber χ^2_M in vielen der folgenden Fitindizes genutzt wird, wird es in den meisten Publikationen mit SEM angegeben (Kline, 2004).

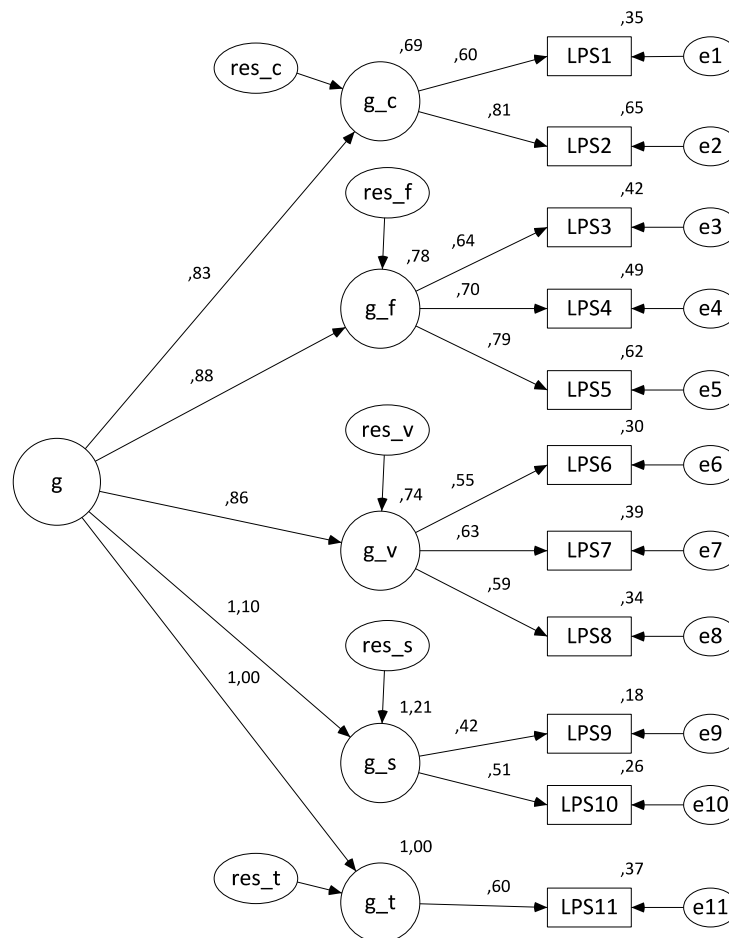


Abbildung 11.1: Ergebnis der Überprüfung des SEM für die Einordnung des LPS-neu in das Drei-Ebenen-Modells von Carroll (1993)

Der Root Mean Square Error of Approximation ($RMSEA$; Steiger, 1990) berücksichtigt die Komplexität des Modells, indem er den Zentralitätsparameter λ berücksichtigt (vgl. Kapitel 7.1) und ins Verhältnis zu den Freiheitsgraden und der Stichprobengröße setzt. Unter Gültigkeit der H_0 ist χ^2_M normal χ^2 -verteilt. Je unwahrscheinlicher H_0 ist, desto non-zentraler die Verteilung und desto größer λ . Nach Boomsma (2000) soll zum $RMSEA$ auch dessen 90 %-Konfidenzintervall angegeben werden. Im vorliegenden Modell liegt der berechnete Wert von .045 mit 90 % Wahr-

scheinlichkeit im Intervall [.021;.066]; somit bleiben alle drei Werte unter den als kritisch angegebenen Grenzen von .08 und [.05;.10] (Kline, 2004). Dieser Fit-Index deutet somit auf die Modellgültigkeit hin.

Der Comparative Fit Index (*CFI*; Bentler, 1990) vergleicht das Modell mit einem Baseline-Modell (independence model), das üblicherweise davon ausgeht, dass die beobachteten Variablen vollständig unabhängig voneinander sind, also die Kovarianzen bzw. Korrelationen zwischen den einzelnen manifesten Variablen null ist. Da $CFI = 1 - \lambda_M/\lambda_B$ umso eher eins ist, je zentraler die Verteilung von χ^2_M im Vergleich zu χ^2_B , sprechen hohe Werte für den Modell-Fit. Üblicherweise sollte $CFI > .90$ sein (Kline, 2004), was für das formulierte Modell der Fall ist ($CFI = .97$). Kline (2004, S. 140) merkt jedoch an, dass ein CFI von eins nicht einen perfekten Modell-Fit bedeutet, sondern lediglich dass $\chi^2_M < df$.

Ein weiteres Maß zur Überprüfung der Goodness-of-Fit ist das Root Mean Square Residual (*RMR*; Anderson & Gerbing, 1984), das über die mittlere Differenz der Elemente der Kovarianzmatrix des Modells und des jeweiligen Elements der empirischen Kovarianzmatrix berechnet wird. Kline (2004) kritisiert, dass das *RMR*, da es auf den Kovarianzen basiert, nicht unabhängig von der Skalierung der beobachteten Variablen ist, und empfiehlt daher das *SRMR* (standardisierter *RMR*) zu nutzen, das die Korrelationen vergleicht. Alternativ können die z-transformierten Variablen eingegeben werden, da ja dann Korrelations- und Kovarianzmatrix identisch sind. Ist der *SRMR*-Wert kleiner als .05, besteht eine kleine Diskrepanz zwischen Modell und Wirklichkeit. Ist der *SRMR*-Wert größer als .08 wird von einer großen Diskrepanz ausgegangen. Werte zwischen .05 und .08 werden als mittlere Diskrepanz bezeichnet (Bollen, 1989). Für das postulierte Modell kann die Diskrepanz als klein angesehen werden ($SRMR = .042$; $RMR = 1.62$).

Als letztes wird von Kline (2004) das Akaike Information Criterion (*AIC*; Akaike, 1987) ausführlicher vorgestellt. Es wird als vorhersagender (predictive) Fit-Index bezeichnet, da es die Likelihood des Modells (also die Wahrscheinlichkeit, dass die vorliegenden Daten aufgrund des postulierten Modells zustande kamen) lediglich um die Komplexität des Modells in Form der Anzahl der Modellparameter korrigiert ($AIC = \chi^2_M + 2q$). Dieses Maß benötigt jedoch immer einen Vergleichsmaßstab, da es absolut nicht interpretiert werden kann. Dennoch gilt, je niedriger der Wert ist, desto zutreffender das Modell. Bei vorliegender Modellüberprüfung beträgt der *AIC* des Modells 114.48; der *AIC* des saturierten Modells (wenn jeder Parameter frei geschätzt wird, also $df = 0$) hingegen 132.00. Der Reduktion der zu schätzenden Parameter folgt somit ein für das Modell sprechender geringerer Anstieg der Likelihood. Analog kann der *AIC* genutzt werden, um verschiedene Modelle zu vergleichen. Kline (2004, S. 144) führt die Formeln und Kurzbeschreibungen einiger weiterer Goodness of Fit Indizes an, die an dieser Stelle unkommentiert bleiben, allerdings in Tabelle 11.2 dargestellt werden, um etwaigen Vertrautheiten mit den diversen Indizes gerecht werden zu können.

Tabelle 11.2: Modell-Fit-Indizes der Einordnung des LPS-neu in das Drei-Ebenen-Modell von Carroll (1993)

χ^2_M	RMSEA	(S)RMR	AIC	BIC	CAIC	GFI	AGFI	PGFI	NFI	RFI	IFI	TLI	PNFI	PCFI	ECVI
62.48 (40, .013)	.045 [.021;.066]	.042, 1.62	114.48	208.33	234.33	.96	.93	.58	.93	.90	.97	.96	.67	.71	.42 [.36;.52]

Insgesamt zeigen alle Indizes einen guten Modell-Fit. Als nächstes sind die Parameterschätzungen zu betrachten und zu beurteilen. McDonalds und Ho (2002) betonen mehrfach, dass alle Parameter mit den zugehörigen Standardfehlern tabelliert werden sollen und die Darstellung des graphischen Modells sparsam und möglichst ohne Angabe der Parameterschätzungen erfolgen sollte. Kline (2004) hingegen favorisiert bei der graphischen Darstellung der Modelle die Angabe der standardisierten Pfadkoeffizienten, also der Schätzungen der quasi bivariaten Korrelationen der verbundenen Variablen bzw. wie im vorliegenden Modell der Faktorladungen. Dies wurde in Abbildung 3.6 übernommen. AMOS hat dabei die Eigenschaft, statt der Fehlervarianzen die durch die manifeste Variable erklärte Varianz anzugeben. Die Differenz zu eins entspricht der nicht geklärten Varianz (Fehlervarianz). In Tabelle 11.3 werden die unstandardisierten Parameter und die Residual- und Fehlervarianzen (sowohl un- als auch standardisiert) angeben.

Tabelle 11.3: Unstandardisierte Regressionsgewichte, Schätzfehler (in Klammern) und Residualvarianz bzw. Fehlervarianz (un- und standardisiert) des Modells

	g	g_c	g_f	g_v	g_s	g_t	Residual- varianz	stand. Resi- dualvar.
g							4.45 (0.93)	.32 (.07)
b_{g_c}	2.30 (0.37)						10.49 (3.52)	.11 (.04)
b_{g_f}	1						1.29 (0.45)	.09 (.03)
b_{g_v}	1.73 (0.28)						4.62 (1.93)	.08 (.03)
b_{g_s}	0.96 (0.16)						-0.69 (0.79)	-.04 (.04)
b_{g_t}	1.99 (0.27)						0	0
b_{LPS1}		1					61.82 (6.39)	.64 (.07)
b_{LPS2}		1.05 (0.13)					19.95 (4.26)	.35 (.07)
b_{LPS3}			1				8.09 (0.83)	.58 (.06)
b_{LPS4}			0.83 (0.09)				4.09 (0.45)	.51 (.06)
b_{LPS5}			1.08 (0.11)				4.02 (0.55)	.37 (.05)
b_{LPS6}				1			41.12 (4.16)	.69 (.07)
b_{LPS7}				0.76 (0.12)			16.13 (1.89)	.61 (.07)
b_{LPS8}				0.77 (0.12)			20.23 (2.17)	.65 (.07)
b_{LPS9}					1		15.36 (1.52)	.82 (.08)
b_{LPS10}					1.99 (0.34)		38.67 (4.46)	.74 (.09)
b_{LPS11}						1	30.62 (3.04)	.63 (.06)

Mit der negativen Residualvarianz von g_s bzw. der damit verbundenen überhöhten Korrelation zwischen g und g_s von 1.10 (vgl. Abbildung 3.6) liegt ein sogenannter Heywood Fall vor. Es resultieren Werte, die faktisch nicht möglich sind, da der ML-Algorithmus alles tut, um das Modell an die Daten anzupassen, unabhängig davon, wie plausibel die Werte sind (Kline, 2004). Wird die Residualvarianz von g_s a priori auf null gesetzt, ergeben sich so gut wie keine Änderungen in der Modellpassung und den geschätzten Parametern. Die Veränderung der Startparameter, wie von Kline (2004, S. 121) empfohlen, bringt in diesem Fall leider nichts. Abgesehen von diesem

einen unschönen Wert, resultieren durchwegs gut interpretierbare Parameterschätzungen. Die Faktorladungen der Stratum II Faktoren auf die beobachteten Variablen (Abbildung 3.6) weisen für die visuelle Wahrnehmung ($g_v \xrightarrow{.55} LPS6$, $g_v \xrightarrow{.63} LPS7$, $g_v \xrightarrow{.59} LPS8$), kristalline ($g_c \xrightarrow{.60} LPS1$, $g_c \xrightarrow{.81} LPS2$) und fluide ($g_f \xrightarrow{.64} LPS3$, $g_f \xrightarrow{.70} LPS4$, $g_f \xrightarrow{.79} LPS5$) Intelligenz sehr gute Werte auf (Cohen, 1988). Die Werte für Verarbeitungsschnelligkeit ($g_s \xrightarrow{.42} LPS9$, $g_s \xrightarrow{.51} LPS10$ bzw. $g_s \xrightarrow{.46} LPS9$, $g_s \xrightarrow{.55} LPS10$, wenn $res_{g_s} = 0$) können ebenso noch als gut bezeichnet werden. Die Korrelation von $LPS11$ mit g_t beträgt .60, wobei die kognitive Geschwindigkeit modellbedingt vollständig durch g determiniert wird. Ebenso „sinkt“ die Korrelation von g auf g_s auf 1, wenn keine Residualstreuung angenommen wird. Die Ladungen von g auf die übrigen Broad Abilities (Carroll, 1993) entsprechen durchaus der erwarteten Höhe, auch wenn dies insofern eingeschränkt werden muss, dass sich der als verhältnismäßig stärker postulierte Zusammenhang zwischen g und g_f (z. B. Gustafsson, 1984) nicht zeigt ($g \xrightarrow{.83} g_c$, $g \xrightarrow{.88} g_f$, $g \xrightarrow{.86} g_v$).

Tabelle 11.4: Diskrepanzen der empirisch geschätzten und der durch das Modell implizierten Korrelationen (unter der Diagonalen) und Kovarianzen

	LPS1	LPS2	LPS3	LPS4	LPS5	LPS6	LPS7	LPS8	LPS9	LPS10	LPS11
LPS1		0.00	-2.36	-2.52	0.90	4.37	-2.77	0.05	-1.43	4.50	3.05
LPS2	.00		-0.78	-0.21	0.36	3.33	-1.49	-1.13	-1.47	2.10	0.78
LPS3	-.07	-.03		0.32	-0.15	-0.38	1.54	2.16	-0.19	0.31	-1.99
LPS4	-.09	-.01	.03		-0.05	0.59	-0.22	0.72	0.19	-0.57	-0.11
LPS5	.03	.01	-.01	-.01		0.70	0.67	0.27	0.63	-2.03	0.27
LPS6	.06	.06	-.01	.03	.03		-0.52	-2.71	-1.54	3.76	0.21
LPS7	-.06	-.04	.08	-.02	.04	-.01		1.73	-0.13	-2.90	-1.52
LPS8	.00	-.03	.10	.05	.02	-.06	.06		-0.49	0.14	-3.79
LPS9	-.03	-.05	-.01	.02	.04	-.05	-.01	-.02		0.00	1.59
LPS10	.06	.04	.01	-.03	-.09	.07	-.08	.00	.00		2.53
LPS11	.05	.02	-.08	-.01	.01	.00	-.04	-.10	.05	.05	

Die höchste Diskrepanz beim Vergleich der Korrelationen der empirischen Daten und der durch das Modell geschätzten Daten beträgt .10 (Tabelle 11.4), was einem Unterschied von 1 % Varianzaufklärung entspricht. Die insgesamt geringen Diskrepanzen spiegeln sich ja bereits im niedrigen und damit adäquaten *SRMR* Fit-Index (vgl. Tabelle 11.2) wieder, bei dessen Interpretation mittlere Werte kleiner .05 als gut erachtet werden, was bei 38 der 55 Korrelationen der Fall ist.

Insgesamt spricht also vieles für einen guten Fit des Modells, wenn der Heywood Fall nicht wäre. Daher soll im Folgenden überprüft werden, ob ein alternatives Modell eine adäquatere Lösung darstellt.

11.4 Modellvergleich

Das Alternativmodell entspricht dem ursprünglichen Modell in den drei Faktoren g_c , g_f und g_v ; anstatt auf g_s und g_t laden jedoch die Subtests 9, 10 und 11 auf einem Faktor, der den Überlegungen von Kapitel 3.2 zufolge als kognitive Geschwindigkeit (g_t) bezeichnet wird (vgl. Abbildung 11.2).

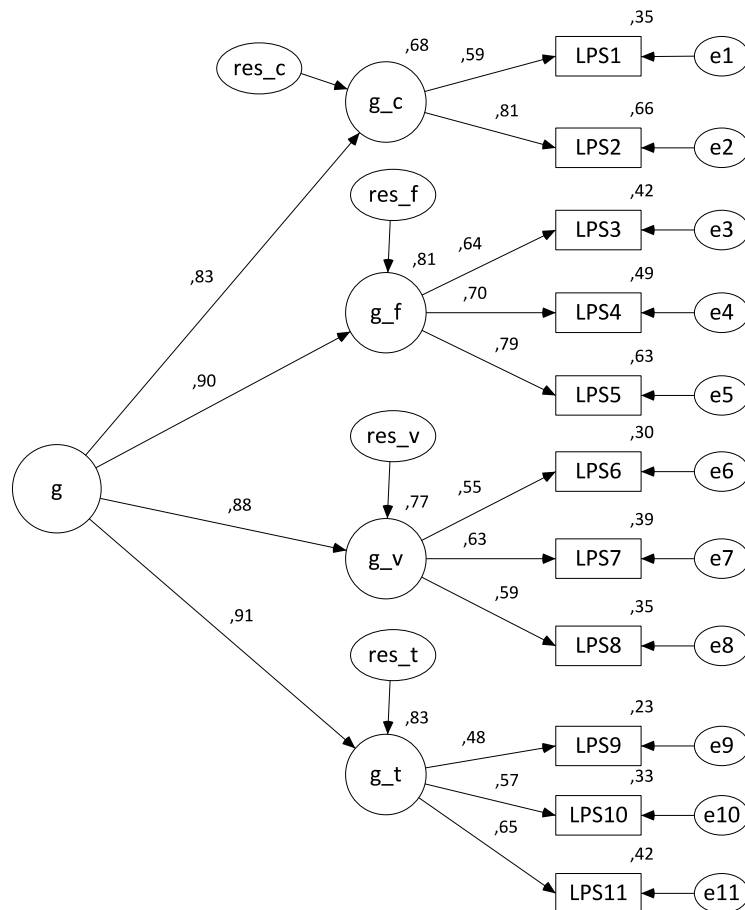


Abbildung 11.2: Ergebnis der Überprüfung des SEM für die Einordnung des LPS-neu in das alternative Modell

Dieses Modell passt den Fit-Indizes nach etwas besser auf die Daten als das ursprüngliche. Insbesondere die geschätzten Parameter stellen nun eine theoretisch mögliche und inhaltlich konsistente und sinnvolle Modellierung dar.

Tabelle 11.5: Modell-Fit-Indizes der Einordnung des LPS-neu in das Alternativmodell

χ^2_M	RMSEA	(S)RMR	AIC	BIC	CAIC	GFI	AGFI	PGFI	NFI	RFI	IFI	TLI	PNFI	PCFI	ECVI
60.39 (40, .020)	.043 [.018;.065]	.041, 1.71	112.39	206.24	232.24	.96	.93	.58	.93	.90	.97	.96	.68	.71	.41 [.35;.51]

Die empirischen Korrelationen der Variablen bleiben natürlich dieselben (vgl. Tabelle 11.4). Die Diskrepanzen der Korrelationen verringern sich augenscheinlich ein wenig; dies wird durch den marginal geringeren SRMR-Wert von .041 (im Vergleich zu .042) nur bedingt untermauert. Die Schätzungen entsprechen sich diesbezüglich ebenso.

Tabelle 11.6: Diskrepanzen der empirisch geschätzten und der durch das alternative Modell implizierten Korrelationen und Kovarianzen, standardisierte Messfehler-varianzen und zugehöriger Standardmessfehler

	LPS1	LPS2	LPS3	LPS4	LPS5	LPS6	LPS7	LPS8	LPS9	LPS10	LPS11	$e_{LPS}(S_e)$
LPS1		0.00	-2.36	-2.52	0.90	4.37	-2.77	0.05	-1.43	4.50	3.05	.65 (.07)
LPS2	.00		-0.78	-0.21	0.36	3.33	-1.49	-1.13	-1.47	2.10	0.78	.34 (.08)
LPS3	-.07	-.04		0.32	-0.15	-0.38	1.54	2.16	-0.19	0.31	-1.99	.58 (.06)
LPS4	-.09	-.02	.03		-0.05	0.59	-0.22	0.72	0.19	-0.57	-0.11	.51 (.06)
LPS5	.03	.01	-.01	-.01		0.70	0.67	0.27	0.63	-2.03	0.27	.37 (.05)
LPS6	.06	.05	-.02	.02	.02		-0.52	-2.71	-1.54	3.76	0.21	.70 (.07)
LPS7	-.06	-.04	.07	-.03	.03	-.01		1.73	-0.13	-2.90	-1.52	.60 (.07)
LPS8	.00	-.03	.09	.04	.00	-.06	.06		-0.49	0.14	-3.79	.65 (.07)
LPS9	-.02	-.02	.00	.03	.06	-.03	.01	-.01		0.00	1.59	.77 (.07)
LPS10	.08	.06	.03	-.01	-.07	.08	-.07	.02	-.06		2.53	.67 (.07)
LPS11	.06	.03	-.07	.00	.02	.01	-.04	-.10	.03	.02		.58 (.07)

Die Fehlervarianzen der manifesten LPS-Variablen ändern sich für die Subtests 1 bis 8 kaum, was auch wegen der unveränderten Messmodelle der jeweiligen Stratum II Faktoren nicht zu erwarten ist. Für die Subtests 9, 10 und 11 hingegen sinken die Fehlervarianzen, zum einen da nun drei manifeste Variablen einer latenten Variable als Indikator dienen, zum anderen da sich diese Zusammenfassung insgesamt als stimmiger und besser zu den Daten passend erweist. Ein weiteres Ergebnis, das die Annahme eines übergeordneten g -Faktors auf Stratum III rechtfertigt, sind die hohen Interkorrelationen der vier Broad Abilities (Tabelle 11.7). Die Residualvarianzen sind insgesamt sehr niedrig, so dass die Modellierung der Stratum II als angemessen und der Struktur der Intelligenzleistung, wie sie durch das LPS-neu erhoben wird, adäquat angenommen werden kann.

Tabelle 11.7: Interkorrelationen der Faktoren 2. und 3. Ordnung des alternativen Modells sowie standardisierte Residualvarianzen und zugehöriger Standardmessfehler

	g	g_c	g_f	g_v	g_t	stand. Residualvar.
g	1					1
g_c	.83	1				.11 (.04)
g_f	.90	.74	1			.08 (.03)
g_v	.88	.72	.79	1		.07 (.03)
g_t	.91	.75	.82	.80	1	.04 (.03)

11.5 Struktur des LPS-neu in der Computerversion

Es erscheint wenig sinnvoll, aufgrund der Interaktionen zwischen Subtests und Bedingungen bzw. der geringen Fallzahlen für jede Bedingungskonstellation die Struktur der Computerversion (CV) des LPS-neu zu überprüfen. Eine Bestätigung des Modells der Papier-Bleistift-Version (PBV) durch die mittels der CV erhobenen Daten über alle Bedingungen hinweg legt hingegen den Schluss nahe, dass durch die unterschiedlichen Einflüsse der Operationalisierung am PC die Kovarianzstruktur nur geringfügig verändert wurde und somit doch prinzipielle Ähnlichkeiten der beiden Erhebungsmethoden bestehen.

Tabelle 11.8: Modell-Fit-Indizes der Einordnung der CV des LPS-neu in das Alternativmodell

χ^2_M	RMSEA	(S)RMR	AIC	BIC	CAIC	GFI	AGFI	PGFI	NFI	RFI	IFI	TLI	PNFI	PCFI	ECVI
65.05 (40, .007)	.053 [.028;.076]	.046, 1.87	117.05	205.64	231.64	.95	.92	.58	.90	.86	.96	.94	.65	.70	.53 [.44;.65]

Wie aus Tabelle 11.8 ersichtlich ist, passen die Ergebnisse der CV der $n = 223$ Berufsschüler, die die CV vollständig bearbeiteten, ähnlich gut zu dem alternativen Modell wie die Ergebnisse der PBV. Die meisten Fit-Indizes deuten eine geringfügig schlechtere Passung an, BIC und CAIC sogar eine marginal bessere. Die Ladungen der Subtests auf die Stratum II Faktoren (Abbildung 11.3) unterscheiden sich mit Ausnahme von Subtest 10 nur gering. Subtest 10 lädt mit .27 (unstandardisiert .49) relativ gering auf g_t ; zudem sind mehr als 90 % Fehlervarianz gegeben. Bei den Zusammenhängen der Stratum II Faktoren zu g fällt lediglich die etwas geringere Ladung von g_c auf (.73 im Vergleich zu .83 in der PBV).

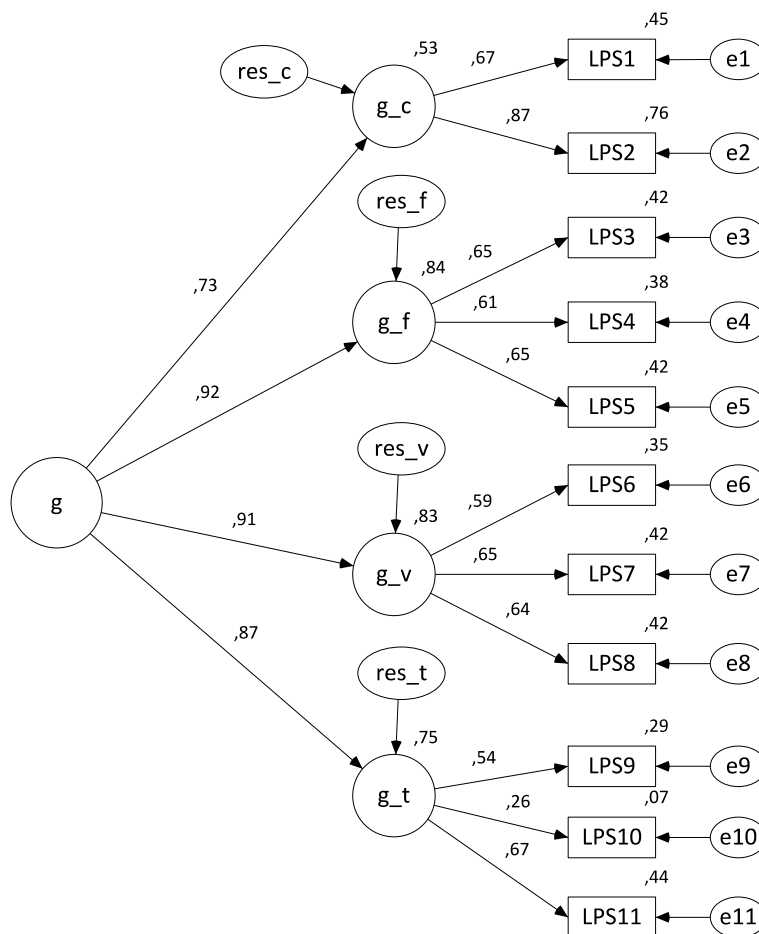


Abbildung 11.3: Ergebnis der Überprüfung des SEM für die Einordnung der CV des LPS-neu in das alternative Modell

Der Vollständigkeit halber sei erwähnt, dass der Modell-Fit der Ergebnisse der CV auf das ursprünglich vorgeschlagene Modell ähnlich leicht schlechtere Werte liefert, der Heywood-Fall an derselben Stelle jedoch noch stärker auftritt.

12. Resümee

12.1 Unberücksichtigte Aspekte

Anhand der vorliegenden Untersuchung hätten noch weitere Fragen ausführlich diskutiert werden können. Im Folgenden werden einige davon kurz aufgegriffen.

12.1.1 Der Einfluss des Geschlechts auf die Testleistung

Die Frage, ob Mädchen oder Jungen bzw. Frauen oder Männer in dem einen oder anderen Leistungsbereich besser abschneiden als die anderen, findet häufig großes Interesse (vgl. Rost, 2009). Unter dem Aspekt der Äquivalenz besitzt die Frage der Interaktion zwischen den Einflüssen der unterschiedlichen Bedingungen der Computerversion (CV) und dem Geschlecht gewisse Relevanz. Ziel für die Konstruktion äquivalenter Verfahren muss jedoch die Unabhängigkeit von Geschlecht und gewählter Darstellung am PC sein, da es nicht zielführend sein kann, für eine prinzipiell gleiche Testung unterschiedliche CV vorzugeben. Bedingt durch die nicht der Grundgesamtheit entsprechende Geschlechterverteilung in der Stichprobe können Geschlechtereffekte nicht valide nachgewiesen werden. Zudem handelt es sich bei den 33 Schülerinnen einer technischen Berufsschule um keine repräsentative Stichprobe weiblicher Probanden, da Ausbildungsinteresse durchaus mit den Stärken und Schwächen kognitiver Fähigkeiten und Schulleistungen kovariiert (Kreuzpointner, 2008; vgl. Kapitel 2.3.2). Der Mittelwertsvergleich der Leistungen von Mädchen und Jungen zeigt, dass die Schülerinnen in den meisten der elf Subtests sowohl in der Papier-Bleistift-Version (PBV) als auch in der CV bessere Leistungen erbracht haben. In acht der 22 univariaten Vergleiche würde der Unterschied sogar statistisch signifikant. Multivariat zeigt sich bei Berücksichtigung der Testwiederholung und der Modusreihenfolge zwar dieser Vorteil der Schülerinnen als Haupteffekt, bedeutsame Interaktionen mit dem Geschlecht liegen jedoch nicht vor. Aufgrund der fehlenden Repräsentativität wird auf den Geschlechtereffekt folglich nicht weiter eingegangen, insbesondere da die Äquivalenz der beiden Modi davon nur sehr bedingt beeinflusst erscheint.

12.1.2 Der Zusammenhang der Testleistung mit der Schulleistung

Es kann als Tatsache bezeichnet werden, dass die Leistung in einem Intelligenztest mit der Schulleistung zusammenhängt und die allgemeine Intelligenz (g) als bester Einzelprädiktor gesehen werden muss (Einen aktuellen Überblick über die Rolle der Intelligenz gibt Rost, 2009.). Eine weitere Tatsache ist, dass die Schulleistung von Berufsschülern unterschiedlicher Ausbildungsrichtungen nicht ohne Weiteres einheitlich und vergleichbar bestimmt werden kann. In einigen Zweigen werden klassische Fächer wie Mathematik unterrichtet, in anderen werden die mathematischen Inhalte in auf die Ausbildung ausgerichtete Fächer eingebunden. Zudem können Berufsschüler keinen, einen Hauptschul-, Realschul- oder gymnasialen Abschluss haben. Vereinzelt besuchen auch Akademiker die Berufsschule, wenn sie nach dem Studium einen Lehrberuf anstreben.

Die Validität der Bestimmung der Schulleistung war somit noch geringer als von Noten gemeinhin angenommen wird. Die Noten der Fächer Deutsch und Englisch korrelieren entsprechend der Erwartung mit den beiden verbalen Subtests 1 und 2 (allerdings nur bedingt hoch $-.20 > r > -.35$). Alle weiteren Korrelationen sind abgesehen von wenigen Ausnahmen zwar entsprechend der Erwartung negativ, jedoch nur einige im Betrag größer .20. Wie sich schon bei der studentischen Stichprobe zeigte, ist der Zusammenhang des LPS-neu mit schulischen Leistungen nicht so stark ausgeprägt wie bei vielen anderen Intelligenztests. Aufgrund der relativ geringen Ausprägung der Zusammenhangsmaße kann die Schulleistung im Kontext der Äquivalenz von PBV und CV nur eine geringe Rolle spielen.

12.1.3 Probabilistische Testtheorie

Es war Teil der Planung dieser Arbeit, die Äquivalenz sowohl im Kontext der Klassischen Testtheorie als auch einer Probabilistischen Testtheorie (PTT) zu überprüfen. Die Vorteile, bei Gültigkeit einer PTT die Äquivalenz getrennt unter den Aspekten der Personenfähigkeit und Aufgabenschwierigkeit zu untersuchen, liegen auf der Hand. Es war jedoch nicht möglich, ein für das relativ komplexe Versuchsdesign passendes Modell zu finden oder die allgemeinsten Modelle (z. B. DeBoeck & Wilson, 2004; Embretson, 1984; Sheng & Wikle, 2007) dahingehend abzuleiten. Die Passung der einzelnen Subtests der PBV für das Raschmodell wurde überprüft und kann nur bedingt als gegeben erachtet werden. Modelle für Tests mit Zeitbegrenzung, wie sie beispielsweise von Verhelst, Verstralen und Jansen (1997) oder bereits von Lord (1980) vorgeschlagen wurden, würden vermutlich die Daten besser erklären können. Zur Überprüfung dieser Modelle sind jedoch sehr große Stichproben und spezielle Programmierungen nötig. Diese Voraussetzung ist für die CV unter dem Aspekt der Dekomposition der Bedingungen Präsentation, Antwortformat und Zeitvorgabe nicht gegeben, so dass die Forschungsfragen hinsichtlich der Testtheorie zurückgestellt wurden. Sie sollten mit einer oder wenigen Bedingungen der CV - ausgewählt basierend auf den Ergebnissen dieser Arbeit - genauer untersucht werden. Damit besteht dann die Möglichkeit, weitere den Sachverhalt hinsichtlich der Äquivalenz komplizierende Aspekte wie beispielsweise die Intelligenzstruktur des LPS zu berücksichtigen. Eine Methode hierzu wurde kürzlich von Cheng, Wang und Ho (2009) vorgestellt.

12.2 Zur Revision des LPS und Anwendbarkeit für die Fragestellung

Die Ergebnisse der psychometrischen Eigenschaften des LPS-neu (Kapitel 10.2) deuten darauf hin, dass auch die Revision des LPS ein hinsichtlich der Gütekriterien und Verteilungen durchaus brauchbares Verfahren ist. Die Analysen der Struktur des LPS-neu im Kontext der drei Ebenentheorie von Carroll (1993; Kapitel 11) zeigen, dass Konstruktvalidität gegeben ist. Weitere Aspekte der Validität sind noch zu untersuchen; sie sind für die Frage der Äquivalenz von CV und PBV nachrangig. Mit dem LPS-neu können durch elf unterschiedliche Subtests vier Faktoren kognitiver Kompetenz zweiter Ordnung (kristalline und fluide Intelligenz, visuelle Wahrnehmung und kognitive Geschwindigkeit) und ein Faktor dritter Ordnung (Allgemeine Intelligenz, *g*) gemessen werden.

12.2.1 Eigenschaften der Papier-Bleistift-Version

Aufgrund der Standardisierungen bei Durchführung, Auswertung und Interpretation können die wichtigsten Aspekte der Objektivität als gegeben erachtet werden. Interne Konsistenz und Split-Half-Reliabilität sind mit Ausnahme der Subtests 3, 4 und 5 durchwegs gut ($> .80$; vgl. Tabelle 10.8). Die Werte von Subtest 3 sind für einen Leistungstest etwas zu niedrig ($\alpha = .76$), was vermutlich an der zu hohen Zahl sehr leichter Items liegt. Dies sollte einfacher zu korrigieren sein als die zu hohe Zahl zu schwerer Items in den Subtests 4 und 5. Bei diesen beiden Subtests müssen vor der endgültigen Revision noch einige Items ausgetauscht und neu konstruiert werden. Die nach der Spearman-Brown-Formel korrigierte Reliabilität von $.74$ für Subtest 4 und $.83$ für Subtest 5 lässt jedoch ohne wesentliche Einschränkung der Gültigkeit Aussagen hinsichtlich der Äquivalenz der beiden Bearbeitungsmodi zu, so dass reliabilitätskorrigierte Schätzungen als unnötig erachtet werden und unterlassen wurden.

Der Vergleich der PBV der beiden Gruppen, die die PBV entweder zu T1 oder zu T2 bearbeiteten, zeigt, dass die zweite Testung bei allen außer Subtest 11 zu deskriptiv besseren Ergebnissen führte; allerdings ist der Unterschied nur in den Subtests 3 und 8 statistisch bedeutsam ($d_3 = 0.58$, $d_8 = 0.36$). Der Zeitgewinn bei den relativ vielen leichten Items, die zu T2 ein zweites Mal bearbeitet wurden, könnte zu dieser Verschiebung der mittleren Leistung geführt haben. Bei Subtest 8 hat zu T2 rund ein Drittel der Probanden alle Items richtig bearbeitet. Eine Verkürzung der Bearbeitungszeit könnte dieses Problem beheben. Für die übrigen Subtests scheint eine Testung am PC einige Wochen vor der Testung mittels Papierbogen nur einen geringen Einfluss auf die Testleistung gehabt zu haben. Die Verteilungen der Stichproben für die einzelnen Subtests (Anhang 15.4) weisen deutliche Ähnlichkeiten in den beiden Modi auf. Schiefe und Exzess fallen zudem nur in wenigen Fällen auf (Tabelle 10.7), so dass insgesamt die Voraussetzungen für die in Kapitel 6 und 7 beschriebenen statischen Verfahren zur Überprüfung der Äquivalenz ebenso wie zur Analyse der Struktur (Kapitel 11) als gegeben erachtet werden können.

12.2.2 Eigenschaften der Computerversion

Die Bestimmung der psychometrischen Eigenschaften der CV ist aufgrund der 15 Durchführungsbedingungen nicht ohne Weiteres möglich (vgl. Kapitel 10.3). Die Betrachtung der mittleren Itemschwierigkeiten und Trennschärfen sowie deren Lageparameter (Tabelle 10.13 und Tabelle 10.14) und der Vergleich mit den Werten der PBV eingeteilt in die entsprechenden Gruppen zeigen jedoch, dass die Subtests in der CV prinzipiell ähnliche Eigenschaften aufweisen. Ebenso entspricht die interne Konsistenz der Subtests in den einzelnen Bedingungen der CV weitestgehend den Ausprägungen der PBV. Bei Vernachlässigung der unterschiedlichen Bedingungen zeigte sich für die CV eine annähernd so gute Passung an das Intelligenzmodell wie für die PBV (Kapitel 11.5). Somit ergeben sich zumindest deutliche Hinweise auf experimentelle Äquivalenz, wie sie von Honacker (1988, vgl. Kapitel 4.3.2) formuliert wurde.

Die in Kapitel 10.5 dargestellten Ergebnisse hinsichtlich des Vergleichs der Bedingungskonstellationen zeigen, dass die Variation der Bedingungen unterschiedliche Effekte begründet. Es ist somit nicht trivial, wie ein Test auf den Computer übersetzt wird. Es wird jedoch auch deutlich, dass die unterschiedlichen Subtests unterschiedlich stark auf Änderungen der Darstellungsform

der Items, der Antwortmodalität und der Zeitvorgabe reagieren. Das LPS-neu und die Umsetzung auf den Computer können somit durchaus als geeignet erachtet werden, die Frage nach Bedingungen der Äquivalenz zu untersuchen. Im Folgenden werden die drei variierten Bedingungen separat diskutiert.

12.3 Bedingungen der Äquivalenz

Bisherige Studien zeigen ein uneinheitliches Bild hinsichtlich der Äquivalenz von PBV und CV (vgl. Kapitel 4.1 und 4.2). In den Studien zur Vergleichbarkeit der Modi des LPS (Frenske, 1992; Klinck, 1998; Troche et al., 2002) waren die Leistungen der verbalen Subtests in der PBV besser, während die Ergebnisse in den anderen Subtests eher in der CV höher waren oder sich entsprachen. Die Korrelationen der Testungen in den beiden Modi waren durchwegs hoch ($> .50$), für parallele Verfahren jedoch meist zu niedrig ($< .80$). Die Metaanalysen (Mead & Drasgow, 1993; Wang et al., 2007) zeigten, dass Äquivalenz für Powertests eher nachgewiesen werden kann als für Speedtests, es aber viele weitere moderierende Einflussfaktoren gibt. Allerdings wurden vorwiegend Stichprobenmerkmale wie das Alter der Probanden und der Rahmen der Testung berücksichtigt und nicht die Merkmale der Umsetzung der CV.

12.3.1 Präsentationsform

Für fünf der elf Subtests (1, 5, 9, 10 und 11) zeigen sich in der CV statistisch bedeutsame Unterschiede zwischen der Einzelitempräsentation und der Gesamtdarstellung (vgl. Tabelle 10.21), wobei in den Subtests 1, 9 und 11 bei der Gesamtdarstellung und in den Subtests 5 und 10 in der Einzelitemdarstellung bessere mittlere Leistungen resultieren (Tabelle 10.20). Für Subtest 9 liegt der Vorteil der Gesamtdarstellung ($\eta_p^2 = .22$) vermutlich in der schnelleren Bearbeitungsmöglichkeit der Subtestaufgabe, die Zeilen zu durchsuchen. Für Subtest 10 dürfte beim Vergleich der beiden Seiten das Fehlen anderer Items von Vorteil sein ($\eta_p^2 = .29$). Bei den Subtests 1 und 11 ist der Unterschied der beiden Darstellungsformen wesentlich geringer ($\eta_p^2 = .02$). Die etwas höhere mittlere Leistung in der Gesamtdarstellung ist vermutlich durch die höhere Anzahl der präsentierten Items begründet, die zu einer höheren Bearbeitungsgeschwindigkeit anregt. Die bessere mittlere Leistung in der Einzelpräsentation bei Subtest 5 ($\eta_p^2 = .02$) könnte durch die bessere Fokussierung auf die einzelnen Items begründet sein. Dann stellt sich jedoch die Frage, weshalb dieser Effekt nicht auch bei Subtest 4 aufgetreten ist, der prinzipiell gleiche und ebenso schwere Items beinhaltet.

Beim Within-Vergleich hinsichtlich der Äquivalenz zwischen PBV und CV weisen bei Berücksichtigung des Modus und des Testzeitpunktes die Subtests 2 und 9 bedeutsame Effekte bezüglich der Präsentationsform auf (je $\eta_p^2 = .02$). In Abbildung 10.11 wird veranschaulicht, dass die einzige Bedingung, die zu gleichen mittleren Leistungen führt, die Bearbeitung von Subtest 9 zu T2 als CV mit Einzeldarstellung ist, während die Einzeldarstellung als CV zu T1 zu deutlich schlechteren Leistungen und die Gesamtdarstellung unabhängig vom Testzeitpunkt zu deutlich besseren Leistungen führt. Für Subtest 2 zeigte sich lediglich eine geringfügig bessere Leistung

der CV zu T1 bei der Einzelpräsentation im Vergleich zur Gesamtdarstellung; für Subtest 2 überwogen somit die Effekte von Modus und Bearbeitungszeitpunkt.

Bei Betrachtung der Vergleiche von PBV und CV der elf Subtests in den 15 Bedingungskonstellationen (vgl. Tabelle 10.24) können die Effektstärken bei Ausschluss der Subtests 9 und 10 sowohl in der Gesamtdarstellung als auch in der Einzeldarstellung bei allen bis auf je eine Bedingung zusammengefasst werden. Jedoch nur bei der Einzeldarstellung bleiben zwei Bedingungen unter einem Effektbetrag von 0.15. Alle Effekte sprechen für eine bessere Leistung in der PBV. Im Mittel zeigt sich für die Einzeldarstellung ein Effekt von -0.35, für die Gesamtdarstellung von -0.30. Wird die Stichprobe nach der Bearbeitungsreihenfolge von CV und PBV differenziert (vgl. Tabelle 10.25 und Tabelle 10.26), weist die Gruppe, die die PBV zu T1 bearbeitete, einen mittleren Effekt von 0.38 in der Einzel- und von 0.46 in der Gesamtdarstellung, und die Gruppe, die die CV zu T1 bearbeitete, einen mittleren Effekt von -0.87 in der Einzel- und von -0.72 in der Gesamtdarstellung auf. Der mittlere Effekt des Vergleichs von PBV und CV zu T1 (Tabelle 10.27) beträgt in der Einzeldarstellung -0.49 und in der Gesamtdarstellung -0.55, zu T2 (Tabelle 10.28) in der Einzeldarstellung 0.11, wobei hier fünf Effekte negativ und vier Effekte positiv sind, und in der Gesamtdarstellung 0.26 (alle Effekte positiv). Insgesamt ist zu berücksichtigen, dass mit Ausnahme der Gesamtbetrachtung nur wenige über die Subtests gemittelte Effektstärken so homogen sind, dass die Zusammenfassung statistisch unkritisch ist.

Dennoch zeigen die Gegenüberstellungen, dass außer beim Vergleich von PBV und CV zu T2 die Abweichungen von Einzel- und Gesamtdarstellung in ähnlichen Größenordnungen liegen. Wird die CV jedoch nach der PBV bearbeitet, tritt der Effekt der Leistungsverbesserung durch eine zweite Testung bei der Gesamtdarstellung konsistent auf, während es bei der Einzeldarstellung in fünf Bedingungen über alle Subtests betrachtet zu geringeren Leistungen in der CV im Vergleich zur PBV kommt.

Die mittleren Korrelationen der Ergebnisse in den beiden Modi liegen sowohl für die Einzel- als auch die Gesamtdarstellung auf einem mittleren Niveau von .51 bzw. .55, wobei die Höhe für die Schlussfolgerung äquivalenter Tests in beiden Fällen zu gering ist.

Ist der Test bzw. die PBV des Tests dem Probanden unbekannt, spielt die Form der Darstellung der Items am PC eine geringe Rolle. Haben die Probanden hingegen den Testbogen der PBV bereits einmal bearbeitet, führt die Ähnlichkeit der Gesamtdarstellung zu geringeren Verzerrungen der mittleren Leistung durch die Bearbeitung am Computer als dies durch die Änderung der Darstellungsform als Einzelitempräsentation geschieht.

12.3.2 Antwortformat

Die drei Antwortformate weisen in den Subtests 5, 9 und 10 bedeutsame Unterschiede auf (Tabelle 10.21). Bei Subtest 5 ist die mittlere Leistung in der Bedingung „weiter“ (nach der Auswahl der Lösung muss ein Button geklickt werden, um zum nächsten Item zu gelangen, bzw. in der Gesamtdarstellung ist das Item zwar noch sichtbar, aber nicht mehr veränderbar) geringer als in den beiden anderen Bedingungen (Tabelle 10.20). Hier könnte somit der Effekt ($\eta_p^2 = .03$), dass es in der Bedingung „einmalig“ (nach der Auswahl der Lösung erscheint automatisch das nächste Item bzw. das beantwortete Item verschwindet in der Gesamtdarstellung) zu größerer Sorgfalt kommt, greifen. In der Bedingung „korrigierbar“ können die Probanden in der Einzel-

präsentation über eine verkleinerte Darstellung der Items am rechten Bildschirmrand zu bereits bearbeiteten Aufgaben zurückkehren, in der Gesamtdarstellung entspricht die CV der PBV. Die Bedingungen „einmalig“ und „korrigierbar“ unterscheiden sich nicht in der erbrachten mittleren Leistung. Bei Subtest 10 ist die mittlere Leistung in der Bedingung „einmalig“ am besten, und bei „weiter“ geringfügig höher als bei „korrigierbar“ ($\eta_p^2 = .04$). Hier könnte der Zeitgewinn durch das Ausbleiben, den „weiter“-Button klicken zu müssen, ursächlich sein. Bei Subtest 9 ergeben sich deutlich geringere mittlere Leistungen für die Bedingung „korrigierbar“ ($\eta_p^2 = .10$), wobei hier die Interaktion mit der Präsentation zu berücksichtigen ist ($\eta_p^2 = .09$). Dieser Effekt ist primär begründet in einer deutlich schlechteren Leistung der Probanden bei der Einzeldarstellung (Abbildung 10.5). In der Gesamtdarstellung unterscheiden sich die Antwortformate nicht. Die Korrekturmöglichkeit über die verkleinerte Darstellung kostete vermutlich ungleich mehr Zeit als sie hinsichtlich der Gesamtzahl richtig gelöster Items Vorteile brachte; oder sie verwirrte die Probanden so sehr, dass dadurch die Leistung beeinträchtigt wurde.

Der Within-Vergleich der Antwortformate der CV mit der PBV zeigt bei Berücksichtigung von Modus und Testreihenfolge nur für Subtest 2 einen bedeutsamen Effekt ($\eta_p^2 = .02$). Probanden, die zu T1 die PBV bearbeiteten, unterscheiden sich dabei nicht in den drei Antwortformaten. Wurde die CV zu T1 bearbeitet, erzielten Probanden in der „weiter“-Bedingung eine etwas bessere mittlere Leistung als Probanden in der Bedingung „einmal“, und diese waren wiederum etwas besser als die Probanden in der Bedingung „korrigierbar“. Dieser Leistungsunterschied vergrößerte sich bei der zweiten Testung als PBV, die für alle drei Gruppen gleich war, so sehr, dass dies die Signifikanz des Effekts begründet. Anders betrachtet zeigen die Probanden der Bedingung „korrigierbar“ die geringste Verbesserung bei der zweiten Testung als PBV und damit am ehesten Äquivalenz. Für die übrigen Subtests scheint das Antwortformat nur geringen Einfluss hinsichtlich der Leistungsunterschiede zu den beiden Testzeitpunkten und in den beiden Modi zu haben.

Bei Betrachtung der Effekte hinsichtlich der mittleren Leistungen von PBV und CV in den 15 Bedingungskonstellationen über beide Erhebungszeitpunkte hinweg (Tabelle 10.24) zeigt sich für das Antwortformat „einmalig“ ein mittlerer Effekt von -0.21, für „weiter“ von -0.31 und für „korrigierbar“ von -0.46. Beim Vergleich der PBV und der CV der Gruppe, die die CV zu T1 bearbeitete, treten deutlich stärkere gemittelte Effekte von -0.75, -0.69 und -0.98 für die drei Antwortformate auf (vgl. Tabelle 10.25). In der Gruppe der Probanden, die die CV zu T2 bearbeitete, zeigen sich deutliche Vorteile für die CV von 0.59, 0.33 und 0.32 (vgl. Tabelle 10.26). Der Vergleich der beiden Modi zu T1 weist unterschiedlich starke Vorteile für die PBV auf (-0.39, -0.53 und -0.62; vgl. Tabelle 10.27), während beim Vergleich der Modi zum T2 positive gemittelte Effekte resultieren (0.13, 0.14, 0.04; vgl. Tabelle 10.28), wobei in jedem Antwortformat sowohl negative als auch positive Werte auftreten. Zudem ist wieder einzuschränken, dass mit Ausnahme des ersten Vergleichs viele Effekte nicht homogen über die Subtests sind.

Es stellt sich somit ein den Einfluss des Antwortformats auf die Äquivalenz betreffend sehr heterogenes Bild dar. Die Leistungsreihenfolge der Probanden in den drei Ausprägungen der Bedingung ist hingegen ähnlich (.61, .54 und .55), jedoch hinsichtlich äquivalenter Verfahren zu niedrig. Beim Antwortformat „einmalig“ zeigen sich die geringsten gemittelten Effekte. Wird die CV zu T2 bearbeitet, erweist sich dieses Antwortformat als besonders vorteilig. Die Probanden profitieren offenbar von dem Zeitgewinn und dem Zwang, ihre Antworten sorgfältig auszuwäh-

len, so dass der generell negative Effekt der Testung am Computer etwas kompensiert wird. Das Antwortformat „korrigierbar“, das als Pendant zur PBV konstruiert wurde, scheint die Probanden hingegen mehr zu verwirren oder aufzuhalten, als dass es Vorteile brächte. Das Format „weiter“, das in sehr vielen Umsetzungen von bestehenden Tests auf den Computer genutzt wird, liegt hinsichtlich seiner Effekte zwischen den beiden anderen Formaten. Eine mögliche Sichtweise wäre, dass hierbei somit die geringste Beeinflussung bzgl. der Äquivalenz vorliegt. Unterschiede im Antwortformat könnten auch die Vorteile der CV des I-S-T 70 in den Studien von Bulheller (1994) und Weibel (1991) erklären, da bei diesem Test die Items in einem Testheft vorgegeben sind und die Lösungen auf einem Antwortbogen markiert werden müssen. Die durch die direkte Eingabe der Lösung am Computer gesparte Zeit könnte dann zu höheren mittleren Leistungen geführt haben.

12.3.3 Zeitvorgabe

Die Darstellung der Bearbeitungszeit erwies sich bei sieben der elf Subtests (2, 3, 5, 7, 9, 10 und 11) als relevante Einflussgröße (Tabelle 10.21). Allerdings hatten die drei Darstellungsformen („ohne“, „gesamt“ und „einzeln“) in unterschiedlichen Subtests unterschiedliche Effekte. Bei Subtest 2 führte das Fehlen der Zeitangabe zu einer geringeren mittleren Leistung ($\eta_p^2 = .03$), insbesondere bei der Gesamtdarstellung der Items. Die Bewusstmachung der Restzeit bzw. der Zwang, sich durch die Begrenzung auf zehn Sekunden pro Item zu beeilen, führte zu besseren Leistungen. Ebenso wiesen die Probanden, die Subtest 7 mit Zeitbegrenzung der Einzelitems bearbeiteten, bessere Leistungen auf ($\eta_p^2 = .04$). Bei Subtest 3 folgten der Beschneidung der Bearbeitungszeit pro Item jedoch geringere Leistungen ($\eta_p^2 = .02$). Bei den Subtests 5 und 11 hatte die Bewusstmachung der Restzeit bessere Ergebnisse zur Folge ($\eta_p^2 = .03$, $\eta_p^2 = .04$), während in den Subtests 9 und 10 das Fehlen dieser Anzeige die besseren Leistungen bedingte, jedoch nur bei Antwortformat „einmalig“ ($\eta_p^2 = .04$, $\eta_p^2 = .10$; vgl. Abbildung 10.2).

Effekte der Zeitdarstellung hinsichtlich des Vergleichs der PBV und der CV bei Berücksichtigung der Bearbeitungsreihenfolge und des Modus weisen dagegen die Subtests 1, 3, 4 und 8 auf (Tabelle 10.23, Abbildung 10.13). Am stärksten traten die Effekte bei Subtest 1 auf. Wird die CV zu T2 bearbeitet, zeigt sich eine deutlichere Verbesserung der mittleren Leistung als in den beiden anderen Bedingungsausprägungen, wenn die Zeitanzeige fehlt. Bei der Bearbeitung der CV zu T1 resultieren hingegen bei der Anzeige der Restzeit höhere Werte als ohne und nochmals niedrigere Werte bei „einzeln“. Insbesondere der letzte Punkt zeigte sich ebenso bei den Subtests 3 und 4. Wenn die Probanden also bereits eine Erfahrung mit der Testung gemacht hatten, scheint das erneute Unwissen um die Bearbeitungszeit zu einer höheren Leistung anregen zu können, während die Begrenzung der Bearbeitungszeit pro Item in diesen Fällen zu äquivalenten Bedingungen hinsichtlich der mittleren Leistung führt, wenn die CV zu T2 bearbeitet wurde.

Über die 15 Durchführungsbedingungen hinweg zeigen die Vergleiche von PBV und CV mit Ausnahme des Vergleichs der Leistungen zu T2 mittlere Effekte. In der Gesamtbetrachtung ist der Unterschied zwischen CV und PBV bei der Anzeige der Restzeit im Mittel am geringsten (-0.26), für die Einzelitemzeitbegrenzung am höchsten (-0.43) und für das Fehlen der Zeitanzeige dazwischen (-.35). Beim Vergleich der Leistungen in CV und PBV der Probanden, die die CV zu T1 bearbeiteten, zeigen sich für alle drei Bedingungen annähernd gleich starke Effekte („ohne“:

-0.84, „gesamt“: -0.78, „einzeln“: -0.80) in Richtung einer besseren Leistung in den PBV. Wird die CV zu T2 bearbeitet, reduziert sich der Vorteil der zweiten Testung in den Bedingungsausprägungen „ohne“ und „gesamt“ etwas (0.56, 0.51), bei der Zeitbegrenzung pro Item hingegen sehr stark (0.11). Werden die Leistungen von CV und PBV der Probanden zu T1 verglichen, ergeben sich deutlich bessere mittlere Leistungen in der PBV („ohne“: -0.47, „gesamt“: -0.48, „einzeln“: -0.67). Beim Vergleich der Ergebnisse zu T2 hat bloß noch das Fehlen der Zeitanzeige einen geringen vorteiligen mittleren Effekt für die CV (0.22), die beiden anderen Bedingungsausprägungen zeigen hingegen keine Effekte mehr (0.00, 0.06), wobei sich auch hier positive und negative Effekte ausgleichen.

Insgesamt zeigt sich, dass die durch eine Begrenzung der Bearbeitungszeit auf 10 Sekunden pro Item intendierte schnellere Bearbeitung der Items nur bei Subtests mit tendenziell sehr leichten Items zu besseren Leistungen führt, dieser Effekt jedoch sehr begrenzt ist. Über alle Subtests hinweg erweist sich diese Bedingungsausprägung eher als erschwerender Faktor, wenngleich er zur höchsten Korrelation zwischen PBV und CV führt (.63). Das Fehlen der Anzeige der Bearbeitungszeit, das der PBV entspricht, erwies sich nicht konsistent als die bessere Bedingung hinsichtlich der Äquivalenz, zeigte aber einen ähnlichen Einfluss wie die Anzeige der Restzeit.

12.3.4 Modus und Testzeitpunkt

Abbildung 10.7 (S. 115) veranschaulicht am deutlichsten die Interaktion zwischen dem Bearbeitungsmodus und dem Bearbeitungszeitpunkt. Für die meisten Subtests (außer 3, 4, 8 und 10) gilt, dass sich die Leistungen der PBV in den beiden Erhebungszeitpunkten kaum unterscheiden (vgl. auch Tabelle 10.6 und Tabelle 10.7)⁸, während die mittleren Leistungen der CV zu T1 deutlich unter und zu T2 eher über den Ergebnissen der PBV liegen. Die Probanden verbesserten sich somit jeweils bei der zweiten Testung, wobei Probanden, die die CV zu T1 bearbeiten, sich meist nur auf das Niveau der Probanden, die die PBV zu T1 bearbeiteten, steigern konnten. Die Leistungssteigerung dieser Probanden fiel hingegen im Mittel deutlich geringer aus.

Bei den Subtests 3, 4 und auch 5 fällt dieser Effekt wesentlich weniger stark auf. Während für Subtest 3 kein Moduseffekt vorliegt, unterschieden sich die mittleren Leistungen in den vier Vergleichsmaßen (PBV zu T1, CV zu T2, CV zu T1, PBV zu T2) der Subtests 4 und 5 generell gering, wobei sich hier die oben beschriebene Interaktion abzeichnet. Auch bei Berücksichtigung der Bedingungsvariationen verschwindet dieser Effekt nicht (vgl. z. B. Abbildung 10.13). Die Reasoning-Aufgaben der Subtests 4 und 5, die insgesamt eine höhere Powerkomponente aufweisen, zeigen sich also weniger anfällig für einen Moduseffekt als leichtere Aufgaben. Diese Feststellung wird durch das Ergebnis von Subtest 10 unterstützt, dessen Items als sehr leicht und prinzipiell von jedem lösbar angesehen werden können. Hier zeigte sich ein sehr deutlicher Moduseffekt dahingehend, dass unabhängig vom Testzeitpunkt im Mittel in der CV weniger Items gelöst wurden als in der PBV.

Auch wenn dieser Effekt in der Bedingung der Einzelitempräsentation deutlich geringer war als bei der Gesamtpräsentation der Items, deuten diese beiden Ergebnisse darauf hin, dass für

⁸ Für den Gesamttestscore der PBV summieren sich die leichten Vorteile der zweiten Testung so stark auf, dass ein bedeutsamer Unterschied zwischen den beiden Gruppen vorliegt ($M_{PBV-CV} = 255$ ($SD = 38$), $M_{CV-PBV} = 268$ ($SD = 41$), $t(271) = 2.79, p < .01$).

die Wahl der Umsetzung eines bestehenden Tests auf den Computer der Aspekt der Bearbeitungsdauer der einzelnen Items eine wesentliche Rolle spielen muss. Die von Franke (1998) zusammengestellten Ergebnisse (vgl. Kapitel 4.1) hinsichtlich der Raven-Matrizen tests unterstützen diese Schlussfolgerung. Häufig wurde übereinstimmend mit dem hohen Poweranteil und der relativ langen Bearbeitungsdauer der Items Äquivalenz berichtet. Wenn sich Unterschiede zeigten, dann lag der Vorteil bei der PBV.

Für die Möglichkeit, den Testwiederholungseffekt unkonfundiert quantifizieren zu können, hätten von zwei weiteren Gruppen Test und Retest im selben Modus erhoben werden müssen. In Anbetracht der vorliegenden Ergebnisse wäre dieses Vorgehen für weitere diesbezügliche Untersuchungen zu empfehlen, sofern der daraus resultierende Mehraufwand einer verdoppelten Stichprobengröße umsetzbar ist. Eine Alternative hierfür bieten die Paralleltests des LPS-neu, die sich während der Fertigstellung dieser Arbeit in der Erprobung befinden. Damit würde sowohl die wiederholte Testung im selben Modus als auch die Testwiederholung im anderen Modus möglich, sofern jeder Proband beide Versionen zu zwei Testzeitpunkten je einmal als CV und einmal als PBV bearbeiten würde.

12.3.5 Fazit

Insgesamt bleibt festzuhalten, dass die Durchführung der Testung am Computer in der Regel zu geringeren Testergebnissen als bei der Testung als PBV führt (vgl. Tabelle 2.24). Abhängig von der Art der Items, insbesondere des Anteils von Power- und Speedkomponenten, und der Form der Antwortabgabe in der PBV (mit oder ohne gesondertem Antwortbogen) können diese Unterschiede durch die geeignete Auswahl der Darstellung der Items und des Antwortformats sowie durch die Anzeige bzw. das Verbergen der Bearbeitungszeit kompensiert werden, was insbesondere bei den Reasoning-Subtests 4 und 5 (vgl. Tabelle 10.33) beobachtbar war. Ist die Bearbeitungszeit pro Item gering, sollten Verzögerungen vermieden werden; ein „weiter“-Button ist in diesem Fall eher kontraproduktiv. Es spricht nur sehr wenig gegen die Einzelpräsentation der Items, so dass die bisherige Praxis diesbezüglich als unkritisch gesehen werden kann und die Chancen der erfolgreichen Umsetzung adaptiver Tests nicht durch dieses Merkmal geschmälert werden. Auch wenn die Anzeige der Restzeit im Vergleich zum Fehlen der Anzeige kaum unterschiedliche Effekte aufweist, soll an dieser Stelle die Empfehlung ausgesprochen werden, die Umsetzung am Computer dem Vorgehen in der PBV anzupassen, da für einige Subtests das permanente Wissen um die verbleibende Bearbeitungsdauer zu ungewollten Effekten führen kann.

Für eine korrekte Übersetzung eines Tests auf den Computer ist die Übereinstimmung von CV und PVB empirisch nachzuweisen. Eine unkontrollierte Übernahme der Normen der PBV wird unter Umständen zu massiven Verzerrungen der Ergebnisse und somit zu gravierenden Fehlern in der Einschätzung der Leistung des Probanden führen.

13. Zusammenfassung

Das Leistungsprüfsystem (LPS; Horn, 1983) ist ein häufig angewandtes Verfahren zur Überprüfung der kognitiven Leistungsfähigkeit. Aufgrund seines fortgeschrittenen Alters ergab sich die Notwendigkeit, über eine Revision nachzudenken. Die Analyse eines großen Teils der Publikationen zum LPS legte eine Neustrukturierung und theoretische Neuorientierung des Verfahrens nahe. Von den bisher 15 Subtests wurden neun überarbeitet übernommen; zwei Neukonstruktionen, basierend auf den Prinzipien alter LPS Subtests, komplettieren die neue Leistungstestbatterie LPS-neu. Die elf Subtests (Allgemeinwissen (1), Anagramme (2), Figurenfolgen (3), Zahlenfolgen (4), Buchstabenfolgen (5), Mentale Rotation (6), Flächenzahl (7), Linienmuster (8), Achtes Zeichen (9), Zeilenvergleich (10), Addieren (11)) messen – bestätigt durch eine konfirmatorische Faktorenanalyse – basierend auf der Drei-Ebenen-Intelligenztheorie von Carroll (1993) vier Faktoren kognitiver Kompetenz zweiter Ordnung (kristalline (1, 2) und fluide (3, 4, 5) Intelligenz, visuelle Wahrnehmung (6, 7, 8) und kognitive Geschwindigkeit (9, 10, 11)) und einen Faktor dritter Ordnung (Allgemeine Intelligenz, g).

Bisher standen nur die ersten vier Subtests des LPS für eine Testung am Computer zur Verfügung. Aufgrund der zunehmenden Relevanz dieser Testform wurde überprüft, wie eine Computerversion (CV) der Subtests des LPS beschaffen sein muss, damit Äquivalenz zwischen der Papier-Bleistift-Version (PBV) und der CV vorliegt. Ziel war jedoch nicht nur, für das LPS-neu äquivalente Bedingungen zu schaffen, sondern generell Verfahrenshinweise zu geben, wie unterschiedlich beschaffene Aufgabenarten und Testzusammenstellungen bei ihrer Übersetzung auf den PC am ehesten äquivalente Ergebnisse für die beiden Modi PBV und CV erzielen können. Hierfür wurden die Darstellung der Items am PC, die Form der Antworteingabe und die Darstellung der Bearbeitungszeit variiert. Zur Überprüfung von Wiederholungseffekten wurde die Reihenfolge der Bearbeitung von PBV und CV variiert (PBV-CV; CV-PBV).

Von 205 Berufsschülern lagen vollständig bearbeitete Tests sowohl der PBV als auch CV des LPS-neu vor (insgesamt nahmen $N = 298$ Schüler an der Untersuchung teil). Die mittleren Leistungen der PBV der Subtests zu den beiden Erhebungen (T1 und T2) unterschieden sich nur für die Subtests 3 und 8 sowie für den Gesamtscore. In den übrigen Subtests führte eine zweite Testung im Mittel lediglich zu deskriptiven Verbesserungen (außer bei Subtest 11, bei dem bei der zweiten Testung im Mittel weniger Items richtig gelöst wurden). Allgemein verbesserten sich die Probanden zwischen der Testung als CV zu T1 zur PBV zu T2 stärker als von der PBV zu T1 zur CV zu T2, woraus ein genereller erschwerender Effekt durch die CV zu folgern ist. Die einzelne Präsentation jedes Items in der CV führt vor allem dann zu Verzerrungen, wenn der Test bereits als PBV bearbeitet wurde. Ist der Test hingegen unbekannt, ist es zweitrangig, ob die Items am PC einzeln (wie häufig üblich) oder alle gleichzeitig (wie in der PBV) angezeigt werden. Insgesamt weist die Präsentationsart der Items bei den meisten Subtests jedoch nur einen geringen Effekt auf. Die Möglichkeit, trotz Einzelitempräsentation über ein Auswahlmenü zu jedem bereits bearbeiteten Item zurückzukehren, führte wider Erwarten zu keiner Erhöhung der Äquivalenz. Auch in der der PBV entsprechenden Darstellung aller Items am PC mit permanenter Korrekturmöglichkeit konnte keine besonders geeignete CV hinsichtlich Äquivalenz gefunden werden. Die mittleren Leistungen werden durch die Anzeige der restlichen Bearbeitungszeit bzw. dem Fehlen der Anzeige nur geringfügig unterschiedlich beeinflusst. Die Begrenzung der Bear-

beitungszeit eines Items auf 10 Sekunden führt in Subtests mit tendenziell leichten Items zu einer Leistungserhöhung; über alle Subtests hinweg reduziert diese Bedingung jedoch die mittlere Leistung.

Für einzelne Bedingungskonstellationen konnte die Äquivalenz einzelner Subtests gezeigt werden. Je wichtiger die möglichst schnelle Bearbeitung der Items in einem Subtest ist, desto sensibler reagieren die Ergebnisse auf die unterschiedlichen Bedingungsvariationen. Bei jeder Übersetzung eines Tests auf den Computer ist die Übereinstimmung von CV und PVB empirisch nachzuweisen und kann nicht a priori angenommen werden.

14. Literaturverzeichnis

- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52, 317–332.
- Amthauer, R. (1955, 1970). *IST Intelligenz-Struktur-Test*. Hogrefe: Göttingen.
- Anastasi, A. (1976). *Psychological testing*. New York: MacMillan Publishing.
- Anderson, J. C. & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, 49 (2), 155-173.
- APA, American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations*. Washington DC: Author.
- Arbuckle, J. L. (2008a). Amos (Version 17) [Computer Program]. Chicago: SPSS.
- Arbuckle, J. L. (2008b). Amos 17 User's Guide. Chicago: SPSS.
- Aurin, K. (1968). Gleiche Chancen im Bildungsgang. In Kultusministerium Baden-Württemberg (Hrsg.), *Bildung in neuer Sicht, Reihe A Nr. 9*. Villingen: Neckar Verlag.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2003). *Multivariate Analysemethoden*. Berlin: Springer.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107 (2), 238-246.
- Bergemann, N., Daumenlang, K. & Johann, G.-K. (1981). Clusteranalytische Befunde zur Intelligenzdifferenzierung. *Psychologische Beiträge*, 23 (3-4), 353-368.
- Biehl, B., Dangel, S. & Reiser, A. (1981). Profile of Mood States. In Collegium Internationale Psychiatriae Scalarum (CIPS, Hrsg.), *Internationale Skalen der Psychiatrie*. Weinheim: Beltz.
- Birnbaum, A. (1968). Some latent trait models. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores* (pp. 397-479). Reading MA: Addison-Wesley.
- Bittner, W. (1960). *Über die Faktorenstruktur der Intelligenz bei Jugendlichen. Untersuchung über die Abhängigkeit der Faktorenstruktur von Neurotizismus*. Unveröffentlichte Hausarbeit Universität Marburg-Lahn.
- Blunch, N. J. (2008). *Introduction to Structural Equation Modelling using SPSS and AMOS*. Sage: Los Angeles.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Boomsma, A. (2000). Reporting on structural equation analyses. *Structural Equation Modeling*, 7 (3), 461-483.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Brandtstädter, J. (1972). Zur empirischen Grundlegung einer Theorie der Beziehungen zwischen „Persönlichkeit“ und „Intelligenz“: Ergebnisse einer Erkundungsstudie. *Psychologische Beiträge*, 14 (4), 38-60.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer.
- Brickenkamp, R. (1972). *Test d2, Aufmerksamkeits- und Belastungs-Test*. Hogrefe: Göttingen.
- Briggs, D. C. & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, 4 (1), 87-100.
- Bulheller, S. (1994). *Untersuchungen zur Äquivalenz ausgewählter Subtests des IntelligenzStruktur-Tests*. Wuppertal: Diplomarbeit im Fach Psychologie, Fachbereich Erziehungswissenschaften.

- Byrne, B. M. (2001). *Structural Equation Modeling with AMOS. Basic Concepts, Applications, and Programming*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Calvert, E. J. & Waterfall, R. C. (1982). A comparison of conventional and automated administration of Raven's Standard Progressive Matrices. *International Journal of Man-Machine-Studies*, 17, 305-310.
- Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, 10 (1), 1-19.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (pp. 69-76). New York: Guilford.
- Casella, G. & Berger, R. L. (2008). *Statistical inference*. Pacific Grove: Duxbury.
- Cattell, R. B. (1952). *Factor analysis. An introduction and manual for the psychologist and social scientist*. New York: Harper and Brothers.
- Cheng, Y.-Y., Wang, W.-C. & Ho, Y.-H. (2009). Multidimensional Rasch analysis of a psychological test with multiple subtests: A statistical solution for the bandwidth-fidelity dilemma. *Educational and Psychological Measurement*, 69 (3), 369-388.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, 33, 107-112.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second Edition). Hillsdale: Erlbaum.
- Conover, W. J. (1999). *Practical nonparametric statistics*. New York: Wiley.
- Collins, M. & Odell, K. (1986). Computerization of a traditional test for nonverbal visual problem solving. *Cognitive Rehabilitation*, 4, 16-18.
- Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Curran, P. J., West, S. G. & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1 (1), 16-29.
- Daniels, J. C. (1971). *Figure reasoning test*. London: Crosby Lockwood & Son.
- DeBoeck, P. & Wilson, M. (2004). Explanatory item response models: A generalized linear and nonlinear approach. New York: Springer.
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2 (3), 292-307.
- Diehl, J. M. & Arbinger, R. (2001). *Einführung in die Inferenzstatistik*. Eschborn: Klotz.
- Düker, H. & Lienert, G. A. (1965). *Konzentrations-Leistungs-Test K-L-T*. Göttingen: Hogrefe.
- Embretson, S. (1984). A general latent trait model for response processes. *Psychometrika*, 49 (2), 175-186.
- Fahrenberg, J., Hampel, R. & Selg, H. (1994). *Das Freiburger Persönlichkeitsinventar FPI. Revidierte Fassung FPI-R und teilweise geänderte Fassung FPI-A1. Handanweisung*. Göttingen: Hogrefe.
- Fenske, B. (1992). *Intelligenzmessung in der Computerdiagnostik*. Unveröffentlichte Diplomarbeit Berlin: Humboldt-Universität, Fachbereich Psychologie.

- Ferguson, G. A. (1941). The factorial interpretation of test difficulty. *Psychometrika* 6 (3), 323-329.
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests. Grundlagen und Anwendungen*. Bern: Huber.
- Fischer, G. H. (1995a). The linear logistic test model. In G. H. Fischer & I. Molenaar (Eds.), *Rasch models - foundations, recent developments, and applications* (pp. 131-155). New York: Springer.
- Fischer, G. H. (1995b). Linear logistic models for change. In G. H. Fischer & I. Molenaar (Eds.), *Rasch models - foundations, recent developments, and applications* (pp. 157-180). New York: Springer.
- Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proceedings of the Royal Society of London, Series A, containing papers of a mathematical and physical character*, 121 (788), 654-673.
- Fleishman, A. I. (1980). Confidence intervals for correlation ratios. *Educational and Psychological Measurement*, 40 (3), 659-670.
- Franke, G. H. (1995). *SCL-90-R. Die Symptom-Checkliste von Derogatis - Deutsche Version*. Göttingen; Beltz Test.
- Franke, G. H. (1998). *Computerunterstützte klinisch-psychodiagnostische Selbstbeurteilungsverfahren im Äquivalenztest*. Lengerich: Pabst-Verlag.
- Gaensslen, H. & Schubö, W. (1973). *Einfache und komplexe statistische Analyse*. München: Reinhardt.
- Gittler, G. (1990). *Dreidimensionaler Würfeltest (3DW)*. Weinheim: Beltz.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5 (10), 3-8.
- Greif, S. (1972). *Gruppenintelligenztests – Untersuchungen am WIT, IST, LPS und AIT*. Frankfurt a. M.: Lang.
- Grissom, R. J. & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah: Lawrence Erlbaum.
- Groffmann, K.-J. & Schneevoigt, I. (1964). Vorläufige Ergebnisse einer Vergleichsuntersuchung an Studenten mit dem Leistungsprüfsystem (LPS) von Horn und dem Intelligenzstrukturtest (IST) von Amthauer. *Schweizerische Zeitschrift für Psychologie und ihre Anwendungen*, 23 (3), 243-252.
- Guilford, J.P. & Hoepfner R. (1976). *Analyse der Intelligenz*. Weinheim: Beltz.
- Gulliksen, H. (1945). The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika*, 10 (2), 79-91.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179-203.
- Hänsgen, K.-D. (1999). *Computereinsatz in der Psychodiagnostik - Stand und mögliche Perspektiven. Forschungsbericht Nr. 141*. Freiburg / Schweiz: Psychologisches Institut der Universität Freiburg.
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6 (2), 107-128.
- Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando: Academic Press.

- Honaker, L. M. (1988). The equivalency of computerized and conventional MMPI administration: A critical review. *Clinical Psychology Review*, 8, 561-577.
- Horowitz, L. M., Strauß, B. & Kordy, H. (1995). *Inventar zur Erfassung Interpersonaler Probleme - Deutsche Version*. Göttingen: Beltz-Test.
- Horn, J. L. (1994). The theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *The encyclopedia of intelligence* (pp. 443—451). New York: Macmillan.
- Horn, J. L. & Blankson, N. (2005). Foundations for better understanding cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment* (pp. 41-68). New York: The Guilford Press.
- Horn, W. (1956). *Begabungstestsystem B-T-S*. Göttingen: Hogrefe.
- Horn, W. (1962, 1983). *Leistungsprüfsystem L-P-S*. Göttingen: Hogrefe.
- Horn, W. (1969). *Prüfsystem für Schul- und Bildungsberatung P-S-B*. Göttingen: Hogrefe.
- Horn, W. neu bearbeitet von H. Lukesch, S. Mayrhofer & A. Kormann (2003). *Prüfsystem für Schul- und Bildungsberatung für 6. bis 13. Klassen - Revidierte Fassung (PSB-R 6-13). Handanweisung*. Göttingen: Hogrefe.
- Horst, P. (1941). A non-graphical method for transforming an arbitrary factor matrix into a simple structure factor matrix. *Psychometrika*, 6, 79-99.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park: Sage.
- Hunter, J. E., Schmidt, F. L. & Jackson, G. B. (1982). *Meta-analysis: cumulating research findings across studies*. Beverly Hills: Sage Publications.
- Irtel, H. (1996). *Entscheidungs- und testtheoretische Grundlagen der Psychologischen Diagnostik*. Frankfurt am Main: Verlag Peter Lang.
- ITC, International Test Commission (2005). *International Guidelines on computer-based and Internet delivered testing*. (Zugriff am 01.01.2010 über <http://www.intestcom.org/guidelines>).
- Jackson, D. N. (1984). *Multidimensional aptitude battery: Manual*. Ontario: Research Psychologist Press.
- Jäger, A. O. (1963). Der Wilde Test, ein neues Intelligenzdiagnostikum. *Zeitschrift für experimentelle und angewandte Psychologie*, 10, 260-278.
- Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195-226.
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test BIS-Test, Form 4*. Göttingen: Hogrefe.
- Kaiser, H. F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.
- Klinck, D. (1998). Papier-Bleistift- versus computer-unterstützte Administration kognitiver Fähigkeitstests: Eine Studie zur Äquivalenzfrage. *Diagnostica*, 44 (2), 61-70.
- Koopmann, W. (1964). *Untersuchung über die Intelligenzstruktur in Abhängigkeit von Alter und Geschlecht*. Unveröffentlichte Vordiplomarbeit, Universität Hamburg.
- Kratzmeier, H. (1976). *Raven-Matrizen-Test. Advanced Progressive Matrices*. Weinheim: Beltz.
- Kreuzpointner, L. (2008). *Statistische Analysen eignungsdiagnostischer Daten: Anwendung multivariater Verfahren im Rahmen einer Untersuchung der Eignungsdiagnostik des Berufsförderungswerks Eckert*. Saarbrücken: VDM Verlag.

- Kreuzpointner, L., Simon, P. & Theis, F. J. (2009). The a_d coefficient as a descriptive measure of the within-group agreement of ratings. *British Journal of Mathematical and Statistical Psychology*, Pre-print article, DOI: 10.1348/000711009X465647.
- Kubinger, K. D. & Farkas, M. G. (1991). Die Brauchbarkeit der Normen von Papier-Bleistift-Tests für die Computer-Vorgabe: Ein Experiment am Beispiel der SPM von Raven als kritischer Beitrag. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 12, 257-266.
- Kučera, M. (1992). *Test koncentrace pozornosti [Test of Concentration of Attention]*. Bratislava: Psychodiagnostika, s.r.o.
- Kury, H. (1972). Vergleichsuntersuchung an Psychologiestudenten mit dem Intelligenzstrukturtest von Amthauer und dem Leistungsprüfsystem von Horn. *Psychologie und Praxis*, 17, 57-67.
- Květon, P., Jelínek, M., Vobořil, D. & Klimusová, H. (2007). Computer-based tests: The impact of test design and problem of equivalency. *Computers in Human Behavior*, 23, 32-51.
- Leary, L. F. & Dorans, N. J. (1985). Implications for altering the context in which test items appear: A historical perspective on an immediate concern. *Review of Educational Research*, 55, 387-413.
- Langfeldt, H.-P. (1975). Ein Beitrag zur Faktorenstruktur des Leistungsprüfsystem (LPS). *Diagnostica*, 21 (3), 123-130.
- Langfeldt-Nagel, M. (1979). *Der Einfluss der Schulbildung auf Intelligenztestleistungen*. Dissertation: Pädagogische Hochschule Rheinland.
- Lee, J. A., Moreno, K. E. & Sympson, J. B. (1986). The effects of test administration on test performance. *Educational and Psychological Measurement*, 46 (2), 467-474.
- Liebert, R. M. & Liebert, L. L. (1995). *Science and behavior: An introduction to methods of psychological research*. Englewood Cliffs, NJ: Prentice-Hall.
- Lienert, G. A. (1966). Die Faktorenstruktur der Intelligenz als Funktion des Neurotizismus. *Zeitschrift für experimentelle und angewandte Psychologie*, 10, 140-159.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse*. Weinheim: Beltz.
- Linn, R. L., Drasgow, F., Camara, W., Crocker, L., Hambleton, R. K., Plake, B. S., Stout, W. & van der Linden, W. J. (2002). Computer-based testing: A research agenda. In C. N. Mills, M. Potenza, J. J. Fremer & W. Ward (Eds.), *Computer-based testing: Building the foundation for future assessments* (pp. 289-299). Hillsdale, NJ: Lawrence Erlbaum.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Erlbaum.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lutz, B. (1983). *Leistungsprüfsystem L-P-S (Dok.-Nr. 0294)*. In Zentrum für Psychologische Information und Dokumentation (ZPID) (Hrsg.), PSYINDEX, Segment Tests (Online-Datenbank). Trier: ZPID.
- Mead, A. D. & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114, 449-458.
- Marks, P., Seeber, A. & Stümper, R. (1975). Untersuchungen zum Geltungsbereich und zur Gültigkeit des Leistungsprüfsystems (LPS) von Horn. *Probleme und Ergebnisse der Psychologie*, 55, 49-64.
- McDonald, R. P. & Ho, M. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7 (1), 64-82.

- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (pp. 136-181). New York: Guilford.
- Meili, R. (1966). *Analytischer Intelligenztest (AIT)*. Bern: Huber.
- Morgan, W. A. (1939). A test for the significance of the difference between the two variances in a sample from a normal bivariate population. *Biometrika*, 31 (1), 13-19.
- Neisser, U. (1967). *Cognitive psychology*. New York: Appleton-Century Crofts.
- Neubauer, A. C., Urban, E. & Malle, B. F. (1991). Ravens Advanced Progressive Matrices: Computergestützte Präsentation versus Standardvorgabe. *Diagnostica*, 37, 204-212.
- Neuman, G. & Baydoun, R. (1998). Computerization of paper-and-pencil tests: When are they equivalent? *Applied Psychological Measurement*, 22 (1), 71-83.
- Norris, M. (1994/1995). Assessing intellectual ability: A comparison between computer based and paper-pencil format using Raven's Standard Progressive Matrices Test. University of Westminster, School of Computer Science & Information System Engineering: *Final Year Project Report BSc (Hons) Cognitive Science*, Supervisor: Brenda Walter.
- Olkin, I., Gleser, L. J. & Derman, C. (1980). *Probability models and applications*. New York: Macmillan Publishing.
- Olson, C. L. (1976). On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 83, 579-586.
- Pearson, K. (1905). Mathematical contributions to the theory of evolution. XIV. On the general theory of skew correlation and non-linear regression. *Drapers' Company Research Memoirs. Biometric Series, II*. Cambridge: Cambridge University Press.
- Pierce, C. A., Block, R. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement*, 64 (6), 916-924.
- Raudenbush, S. W. & Bryk, A. S. (2006). *Hierarchical linear models*. Thousand Oaks: Sage.
- Rasch, G. (1960; 1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: Univ. of Chicago Press.
- Raven, J. C. (1958). *Standard Progressive Matrices*. London: H. K. Lewis.
- Raven, J. C., Court, J. H. & Court, J. (1977). *Manual for Raven's progressive matrices and vocabulary scales*. London: H. K. Lewis.
- Reinert, G., Baltes, P. B., Schmidt, L. R. (1965). Faktorenanalytische Untersuchungen zur Differenzierungshypothese der Intelligenz: Die Leistungsdifferenzierungshypothese. *Psychologische Forschung*, 28, 246-300.
- Richardson, M. W. (1936). The relation between the difficulty and the differential validity of a test. *Psychometrika*, 1 (2), 33-49.
- Rijmen, F. & Briggs, D. C. (2004). Multiple person dimensions and latent item predictors. In P. De Boeck and M. Wilson (Eds.), *Explanatory item response models: A generalized linear and non-linear approach*, Springer. New York.
- Rock, D. L. & Nolen, P. A. (1982). Comparison of the standard and computerized versions of the Raven coloured progressive matrices test. *Perceptual and Motor Skills*, 54, 40-42.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills: Sage.
- Rost, D. H. (2009). *Intelligenz. Fakten und Mythen*. Weinheim: Beltz.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.

- Rustenbach, S. J. (2003). *Metaanalyse: Eine anwendungsorientierte Einführung*. Bern: Huber.
- Schumacher, E. (1988). *Selbstbild, Arbeits- und Leistungsverhalten bei Arbeitslosen im Vergleich zu Erwerbstätigen – eine empirische Untersuchung*. München: Profil.
- Shepard, R. N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701-703.
- Schoenfeldt, L. F. (1989). Guidelines for computer-based psychological tests and interpretations. *Computers in Human Behavior*, 5 (1), 13-21.
- Science Research Associates (1984). *Office Skills Tests (OST): Examiner's manual*. Chicago: Author.
- Sheng, Y. & Wikle, C. K. (2007) Comparing multiunidimensional and unidimensional item response theory models, *Educational and Psychological Measurement*, 67 (6), 899-919.
- Schmid, J. & Leimann, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 53-61.
- Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Cambridge: Hogrefe & Huber.
- Smithson, M. J. (2003). *Confidence intervals, quantitative applications in the social sciences series, No. 140*. Thousand Oaks: Sage.
- Senka, J., Kuruc, J. & Čečer, M. (1992). *Bourdonov test [Bourdon's Test]*. Bratislava: Psychodiagnostika, s.r.o.
- Spearman, C. & Jones, L. W. (1950). *Human Ability: A Continuation of "The Abilities of Man"*. London: Macmillan & Company.
- Steck, P., Rausche, A. & Steege, F. W. (1975). Faktorenanalytischer Vergleich gebräuchlicher deutschsprachiger Intelligenztests. *Psychologie und Praxis*, 19, 145-158
- Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, 9 (2), 164-182.
- Steiger, J. H. & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221-257). Mahwah: Erlbaum.
- Steller, M. & Stürmer, M. (1984). Space. Reasoning oder was? - Zur Validität des "Raumvorstellungstests" LPS 8/PSB 7. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 31 (4), 671-692.
- Steyer, R. & Eid, M. (2001). *Messen und Testen*. Berlin: Springer.
- Sturm, W., Willmes, K. & Horn, W. (1993). *Leistungsprüfsystem für 50-90jährige (LPS 50+)*. Göttingen: Hogrefe.
- Sykes, R. C. & Ito, K. (1997). The effects of computer administration on scores and item-parameter estimates of an IRT-based licensure examination. *Applied Psychological Measurement*, 21 (1), 51-63.
- Tent, L. (1969). *Die Auslese von Schülern für weiterführende Schulen*. Göttingen: Hogrefe.
- Thurstone, L. L. (1938, repr. 1969). *Primary Mental Abilities*. Chicago: University of Chicago Press.
- Thurstone, L. L. (1944). Second-order factors. *Psychometrika*, 9 (2), 71-100.
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- Thurstone, L. L., & Thurstone, T. G. (1941). *Factorial studies of intelligence*. Chicago: University of Chicago Press.

- Troche, S., Rammstedt, B. & Rammsayer T. (2002). Vergleich einer Papier-Bleistift- und einer computergestützten Version des Leistungsprüfsystems (LPS). *Diagnostica*, 48, 115-120.
- Überla, K. (1971). *Faktorenanalyse*. Berlin: Springer.
- Verhelst, N. D., Verstralen, H. H. V. M. & Jansen, M. G. H. (1997). A logistic model for time-limit tests. In W. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 169-186). New York: Springer.
- Wang, S., Jiao, H., Young, M. J., Brooks, T. E. & Olson, J. (2007). A meta-analysis of testing mode effects in Grade K-12 mathematics tests. *Educational and Psychological Measurement*, 67, 219-238.
- Welch, B. L. (1949). Further note on Mrs. Aspin's tables and on certain approximations to the tabled function. *Biometrika*, 36, 293-296. [als Appendix zu Aspin, A. A. (1949). Tables for use in comparisons whose accuracy involves two variances, separately estimated. *Biometrika*, 36, 290-296.]
- Weibel, H. L. (1991). *Computer-gestützte Diagnostik in der Schulpsychologie: Anfang einer neuen diagnostischen Ära oder Bedrohung seriöser diagnostischer Tätigkeit?* Dissertation, Universität Zürich.
- Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W. & Delazer, M. (2003). Sex differences in cognitive functions. *Personality and Individual Differences*, 35, 863-875.
- Weiss, R. (1972). Zusammenhänge zwischen Schulleistung und Intelligenz, Bericht über eine Untersuchung an 4000 Vierzehnjährigen in Oberösterreich. *Erziehung und Unterricht*, 122, 663-679.
- Wewetzer, K.-H. (1968). Die Faktorenstruktur der Intelligenz in ihrer Abhängigkeit von der Begabungsart. Ein Beitrag zur Typenanalyse. In K. J. Groffmann & K. H. Wewetzer (Hrsg.). *Person als Prozess. Festschrift zum 65. Geburtstag von Prof. Dr. phil Robert Heiss* (S. 213-230). Bern: Huber.
- Wilks, S. S. (1946). Sample Criteria for Testing Equality of Means, Equality of Variances, and Equality of Covariances in a Normal Multivariate Distribution. *The Annals of Mathematical Statistics*, 17 (3), 257-281.
- Wishart, J. (1932). A note on the distribution of the correlation ratio. *Biometrika*, 24 (3-4). 441-456.
- Zimmerman, D. W. & Williams, R. H. (1997). Properties of the Spearman correction for attenuation for normal and realistic non-normal distributions. *Applied Psychological Measurement*, 21 (3), 253-270.

15. Anhang

15.1 Testbögen des neuen LPS









































<p>Code: _____ (Vater - Mutter - Geb., z.B. FR-CH-12)</p> <p>Bitte jeweils die ersten beiden Buchstaben vom Vornamen von Vater und Mutter sowie Ihren Geburtstag angeben.</p> <p>Geburtstag: _____. _____. _____. Geschlecht: o weibl. / o männl. Datum: _____. _____. _____. Wie fühlen Sie sich? o o o o o o (vor) sehr gut - gut - geht so - schlecht - sehr schlecht o o o o o o (nach)</p>	<p style="text-align: center;">①</p> <p>Markieren Sie den falschen Buchstaben.</p> <p style="text-align: center;">K R ✕ I D E</p> <p>Um eine falsche Auswahl aufzuheben, zeichnen Sie einen Kreis darum und kreuzen Sie stattdessen ihre alternative Lösung an.</p> <p style="text-align: center;">T E ✕ L ✕ R</p>	<p style="text-align: center;">②</p> <p>Markieren Sie den Anfangsbuchstaben.</p> <p style="text-align: center;">G ✕ W E R</p> <p>2. Beispiel:</p> <p style="text-align: center;">C K E R ✕</p>
<p style="text-align: center;">③</p> <p>Markieren Sie die Form, die nicht in die Reihe passt.</p> <p style="text-align: center;">○ ○ ○ ○ ✕ ○ ○ ○</p> <p>Es passt immer nur eine Form nicht in die Reihe!</p> <p style="text-align: center;"> + + + ✕</p>	<p style="text-align: center;">④</p> <p>Markieren Sie die Zahl, die nicht in die Reihe passt.</p> <p style="text-align: center;">2 2 2 2 2 ✕ 2 2 2</p> <p>Es passt immer nur eine Zahl nicht in die Reihe!</p> <p style="text-align: center;">4 ✕ 5 4 4 5 4 4 5</p>	<p style="text-align: center;">⑤</p> <p>Markieren Sie den Buchstaben bzw. das Buchstabenpaar, das nicht in die Reihe passt.</p> <p style="text-align: center;">a b a b a b a ✕ a</p> <p>Es passt immer nur ein Buchstabe bzw. ein Buchstabenpaar nicht in die Reihe!</p> <p style="text-align: center;">aa ✕ b aa aa b aa aa b</p>
<p style="text-align: center;">⑥</p> <p>Markieren Sie das Zeichen, das spiegelverkehrt abgebildet ist.</p> <p style="text-align: center;">\ / \ ✕ /</p> <p>2. Beispiel:</p> <p style="text-align: center;">P ✕ P P P</p>	<p style="text-align: center;">⑦</p> <p>Wie viele Flächen hat der abgebildete Körper?</p> <div style="text-align: center;"> </div> <p>2. Beispiel:</p> <div style="text-align: center;"> </div>	<p style="text-align: center;">⑧</p> <p>Welche Form passt in das Linienmuster?</p> <div style="text-align: center;"> </div> <p>Zwischenlinien sind unerheblich!</p> <div style="text-align: center;"> </div>
<p style="text-align: center;">⑨</p> <p>Markieren Sie jede achte 0. ACHTUNG: nur in Spalte 9</p> <p style="text-align: center;">A00B00 00L0M✕ 0S000R 00F0✕T</p> <p>Wenn Sie am Ende der Spalte angelangt sind, beginnen Sie, jede achte 1 zu zählen, dann jede achte 2, jeder achte 3 usw.</p>	<p style="text-align: center;">⑩</p> <p>Markieren Sie die Zeichen in Spalte 10, die nicht mit den Zeichen in Spalte 9 übereinstimmen.</p> <p style="text-align: center;">A 0 0 B 0 0 A 0 ✕ B 0 0</p> <p>Es gibt auch Zeilen, die vollständig identisch sind, und Zeilen, die sich in mehr als einem Zeichen unterscheiden!</p> <p style="text-align: center;">0 0 L 0 M 0 0 0 L 0 ✕ 0</p>	<p style="text-align: center;">⑪</p> <p>Addieren Sie die zehn Zahlen und markieren Sie die Einerziffer der Lösung.</p> <p style="text-align: center;">2 4 2 6 2 4 2 6 2 ✕</p> <p>Die Einerziffer der Summe ist unter den zehn Zahlen immer nur einmal aufgeführt!</p> <p style="text-align: center;">8 2 6 ✕ 4 2 6 4 2 6</p>

[illegible]











5

6

7

abcdefghijklmn opqrstuvwxyz																						
a	b	a	b	a	b	a	a	a	a	↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
a	b	a	b	a	b	a	a	a		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
a	b	a	b	a	b	a	a	a		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
a	b	a	b	a	b	a	a	a		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
a	b	a	b	a	b	a	a	a		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
a	b	a	b	a	b	a	a	a		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
a	b	a	b	a	b	a	a	a		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
a	b	a	b	a	b	a	a	a		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
a	b	a	b	a	b	a	a	a		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
a	b	a	b	a	b	a	a	a		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
a	b	a	b	a	b	a	a	a		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10
aa	aa	a	aa	aa	b	aa	aa	b		↖	7	↖	↖	7		2 5 8	3 6 9	4 7 10		2 5 8	3 6 9	4 7 10

11

<div>↓</div> <div>↓</div>																																																																																																																							
---------------------------	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	--

15.2 Instruktion des neuen LPS

Allgemeines

In der folgenden etwa 60-minütigen Untersuchung soll festgestellt werden, welche Aufgabenarten dem einzelnen leicht fallen und was ihm weniger liegt. Die gesamte Untersuchung ist daher in elf Tests unterteilt. Die Zeit ist meist zu kurz, als dass es möglich wäre, alle Aufgaben eines Tests zu bearbeiten. Selbst der, der ungewöhnlich schnell arbeitet, wird selten zur Bearbeitung der letzten Aufgaben kommen. Es ist jedoch wichtig, dass Sie sich immer Mühe gibt, damit Sie nicht falsch beurteilt werden. Bitte benutzen Sie einen Stift, der deutlich sichtbar ist.

[Während des Austeilens kann schon mit der Ansage zum ausfüllen des kleinen Feld links oben auf der ersten Seite begonnen werden.]

Bitte lassen Sie den Bogen geschlossen auf der ausgeteilten Seite liegen, so dass sie links oben ein Feld sehen, in dem Code steht. Bitte tragen Sie dort Ihren Versuchspersonencode ein: Dieser besteht aus den ersten beiden Buchstaben des Vornamens Ihres Vater, den ersten beiden Buchstaben des Vornamens Ihrer Mutter und der Tageszahl Ihres Geburtstages. Anschließend füllen Sie bitte die restlichen Felder aus. Das heutige Datum ist der ... Legen Sie anschließend Ihren Stift wieder weg. Drehen Sie den Bogen bitte erst um, wenn ich Sie dazu auffordere. Die Aufgaben im Inneren sind Seitenverkehrt abgedruckt. Man muss das Blatt also umdrehen.

1. Wir kommen nun zu Test 1 (die Instruktion oben in der Mitte):

In diesem Test sehen Sie eine Liste von Wörtern, die jeweils einen Druckfehler enthalten. Genau ein Buchstabe wurde durch einen falschen ersetzt. Ihre Aufgabe ist es, den falschen Buchstaben durch ein Kreuz zu kennzeichnen. Bei KRAIDE ist das A falsch, da KREIDE mit E geschrieben werden muss. Im zweiten Beispiel TELLOR wurde versehentlich das L angekreuzt. Um eine falsche Auswahl aufzuheben, zeichnen Sie einen Kreis darum und kreuzen Sie stattdessen ihre alternative Lösung an. Im Beispiel wurde das O markiert, da TELLER das gesuchte Wort ist. Es zählen nur Aufgaben mit einem **einzigsten eindeutigen** Kreuz. Wenn Sie das Blatt **nach** meiner Aufforderung gleich aufschlagen werden, finden Sie ganz links eine Reihe mit 60 Wörtern. Bitte bearbeiten Sie die Reihe von oben nach unten. Arbeiten Sie so schnell, aber auch so sorgfältig wie möglich. Halten Sie sich mit einem Wort nicht zu lange auf, lassen Sie aber auch keines aus. Im Zweifel raten Sie bitte. Sie haben für die gesamte Aufgabe 3 Minuten Zeit. Falls Sie in dieser Zeit bis an das Ende der Liste gelangen sollten, kontrollieren Sie bitte ihre Antworten nochmals. Beginnen Sie nicht mit dem nächsten Test!

Gib es noch Fragen? [Bei Fragen bitte nur Inhalte der bereits getätigten Instruktion wiederholen und keine weiteren Hilfen geben!]

Dann schlagen Sie das Arbeitsblatt bitte auf, drehen Sie das Blatt um und beginnen Sie jetzt links oben unter der 1. [3 Minute ab da an]

Nach 3 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionssseite zurück.“

2. Schauen wir uns nun die Beschreibung von Test 2 rechts oben an.

In diesem Test geht es um Ihnen prinzipiell bekannte Worte, deren Buchstaben durcheinander angeordnet sind. Zudem wurden die Buchstaben verstümmelt. Ziel bei diesen Aufgaben ist es, den Anfangsbuchstaben des Wortes, das sich aus **genau diesen fünf** Buchstaben bilden lässt, anzukreuzen. Aus den Buchstaben G-Z-W-E-R lässt sich das Wort ZWERG bilden, daher muss das Z markiert werden. Aus den Buchstaben des zweiten Beispiels C-K-E-R-A lässt sich das Wort ACKER bilden. Daher ist das A die richtige Lösung. Auch hier gilt wieder: Es zählen nur Aufgaben mit einem einzigen eindeutigen Kreuz. Versehentlich gemachte Kreuze können durch einen Kreis rückgängig gemacht werden. Arbeiten Sie so schnell, aber auch so sorgfältig wie möglich, lassen Sie bitte kein Wort aus, halten Sie sich aber mit einem Wort auch nicht zu lange auf.

Fragen?

Für diese Aufgabe haben Sie drei Minuten Zeit. Bitte schlagen Sie das Arbeitsblatt auf und beginnen Sie jetzt. [Ab nun **3 Minuten**].

Nach 3 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionssseite zurück.“

3. Betrachten wir nun die Angaben zu Test 3.

Bei diesem Test werden Ihnen pro Aufgabe acht Symbole präsentiert, deren Anordnung und Form einer Gesetzmäßigkeit unterliegen. Ihre Aufgabe ist es, **das** Symbol zu markieren, das dieser Systematik zuwider läuft, also nicht in die Reihe passt. Im Beispiel passt der kleine Kreis nicht zu den sieben großen Kreisen. Das zweite Beispiel folgt der Regel | + | + usw. Daher ist der letzte Strich anzukreuzen, da dort ein + stehen müsste. Auch hier gilt wieder, es zählen nur Aufgaben mit einem einzigen eindeutigen Kreuz. Versehentlich gemachte Kreuze können durch einen Kreis rückgängig gemacht werden. Es gibt immer nur **ein** Zeichen, das nicht in die Systematik passt. Falls Sie der Meinung sind, mehrere Zeichen passen nicht, haben Sie vermutlich einen Aspekt der Systematik nicht beachtet. Versuchen Sie, möglichst jede Aufgabe zu lösen. Verbleiben Sie aber auch nicht zu lange bei einer Aufgabe. Wenn Sie partout keine Lösung finden, raten Sie.

Fragen?

Für diese Aufgabe haben Sie drei Minuten Zeit. Bitte schlagen Sie das Arbeitsblatt auf und beginnen Sie jetzt. [Ab nun **3 Minuten**].

Nach 3 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionssseite zurück.“

4. Wir kommen nun zu Test 4.

Jeder Aufgabe besteht aus neun Zahlen, deren Anordnung und Ausprägung einer Gesetzmäßigkeit unterliegen. Ihre Aufgabe ist es, die Zahl anzukreuzen, die dieser Systematik zuwider läuft, also nicht in die Reihe passt. Im ersten Beispiel passt die 3 nicht zu den acht 2 und muss daher markiert werden. Die richtige Systematik im zweiten Beispiel ist 4 4 5 4 4 5 4 4 5. Die 5 an der zweiten Stelle passt nicht und muss daher markiert werden. Es kann immer nur **eine Zahl falsch** sein, und folglich werden nur solche Aufgaben gewertet, die nur ein eindeutiges Kreuz haben. Auch hier müssen Sie nach einem weiteren Aspekt der Systematik suchen, wenn Sie mehrere Zahlen als nicht passend erachten. Noch mal: Versehentlich gemachte Kreuze können durch einen Kreis berichtigt werden. Die Aufgaben werden etwas schneller schwerer, als in den Tests zuvor. Versuchen Sie dennoch, für jede Reihe eine Lösung zu finden und raten Sie erst, wenn Sie wirklich keine Chance sehen, die Aufgabe zu lösen.

Fragen?

Für diese Aufgabe haben Sie fünf Minuten Zeit. Bitte schlagen Sie das Arbeitsblatt auf und beginnen Sie jetzt. [Ab nun 5 Minuten].

Nach 5 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionssseite zurück.“

5. In Test 5 ...

... werden Ihnen immer neun Buchstaben präsentiert, deren Anordnung sowie Groß- und Kleinschreibung einer Gesetzmäßigkeit unterliegen. Ihre Aufgabe ist es nun, den Buchstaben bzw. das Buchstabenpaar anzukreuzen, der bzw. das dieser Systematik zuwider läuft, also nicht in die Reihe passt. Im ersten Beispiel passt das vorletzte **a** nicht in die Folge a b a b und muss daher markiert werden. Die richtige Systematik im zweiten Beispiel ist aa aa b aa aa b aa aa b. Das dritte, allein stehende **a** passt folglich nicht und muss markiert werden. Auch hier kann **immer nur eine Position falsch** sein. Auch diese Aufgaben werden etwas schneller schwerer. Versuchen Sie auch hier, für jede Reihe eine Lösung zu finden und raten Sie erst, wenn Sie wirklich keine Chance sehen, die Aufgabe zu lösen. Zur Unterstützung ist am Anfang der Spalte ein Alphabet abgebildet.

Fragen?

Für diese Aufgabe haben Sie fünf Minuten Zeit. Bitte schlagen Sie das Arbeitsblatt auf und beginnen Sie jetzt. [Ab nun 5 Minuten].

Nach 5 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionssseite zurück.“

6. Kommen wir nun zu Test 6.

Hierbei werden pro Zeile, also pro Aufgabe fünf gleiche Zeichen (Zahlen oder Buchstaben) abgebildet, die in ihrer Lage verdreht sind. Ein Zeichen ist zusätzlich spiegelverkehrt gedruckt. Ihre Aufgabe ist es, dieses spiegelverkehrte Zeichen anzukreuzen. Richtet man im ersten Beispiel alle 1 wieder ordnungsgemäß auf, sieht man, dass die vierte 1 spiegelverkehrt abgebildet und somit zu markieren ist. Im zweiten Beispiel ist das zweite P gespiegelt und somit zu markieren. Bitte lösen Sie die Aufgaben der Reihe nach von oben nach unten und lassen Sie keine Aufgabe aus. Wenn Sie innerhalb der Zeit fertig werden, kontrollieren Sie Ihre Lösungen nochmals von oben nach unten.

Fragen?

Für diese Aufgabe haben Sie zwei Minuten Zeit. Bitte schlagen Sie das Arbeitsblatt auf und beginnen Sie jetzt. [Ab nun **2 Minuten**].

Nach 2 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionsseite zurück.“

7. Bei Test 7 ...

... werden Ihnen räumlich gezeichnete Körper präsentiert. Ihre Aufgabe ist es, die Zahl der Flächen zu ermitteln und rechts neben dem Körper anzukreuzen. Der Quader im ersten Beispiel hat sechs Flächen, drei sichtbare und entsprechend seiner Form drei nicht-sichtbare. Der Keil im zweiten Beispiel hat fünf Flächen. Daher ist das Kreuz auf der 5. Eine Fläche kann gerade oder gewölbt sein. Nur bei Kanten bzw. Knicken beginnt eine neue Fläche. Vergessen Sie nicht, die nicht sichtbaren Flächen mitzuzählen. Nicht-sichtbare Flächen entsprechen immer der zu erwartenden Form, so dass immer nur eine Zahl richtig sein kann. Folglich werden nur solche Aufgaben gewertet, die ein eindeutiges Kreuz haben. Versehentlich gemachte Kreuze können durch einen Kreis berichtigt werden. Beginnen Sie nach dem Kommando bitte mit der Aufgabe links oben und lösen die Aufgaben der linken Spalte von oben nach unten. Anschließend bearbeiten Sie bitte die rechte Spalte. Bitte halten Sie sich an die Reihenfolge und bearbeiten Sie die Aufgaben nicht durcheinander.

Fragen?

Für diese Aufgabe haben Sie drei Minuten Zeit. Bitte schlagen Sie das Arbeitsblatt auf und beginnen Sie jetzt. [Ab nun **3 Minuten**].

Nach 3 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionsseite zurück.“

8. Betrachten wir das Beispiel von Test 8.

In diesem Test werden Ihnen Linienmuster präsentiert, in welchen sich jeweils genau eine der fünf Formen rechts daneben einbetten lässt. Sowohl Form als auch Lage müssen passen. Im ersten Beispiel passt das L-förmige Linienmuster. Im zweiten Beispiel passt nur das U-förmige Muster, da sowohl das L- als auch das T-förmige nicht vollständig deckungsgleich dargestellt sind. Die Striche innerhalb des U-förmigen Musters sind unerheblich, das U lässt sich ja dennoch deckungsgleich einbetten. Es passt immer nur eine der fünf Formen in der nebenan dargestellten Lage. Versehentlich gemachte Kreuze können durch einen Kreis berichtigt werden. Auch hier sind die Aufgaben in zwei Reihen angeordnet. Bearbeiten Sie zuerst die linke Spalte von oben nach unten und anschließend die rechte Spalte – ebenfalls von oben nach unten. Bitte arbeiten Sie weiterhin konzentriert und so schnell aber auch so sorgfältig wie möglich. Wenn Sie innerhalb der Zeit fertig werden, kontrollieren Sie wieder von Anfang an Ihre Lösungen, also zuerst die linke Spalte und anschließend die rechte.

Fragen?

Für diese Aufgabe haben Sie drei Minuten Zeit. Bitte drehen Sie das Arbeitsblatt nun um und beginnen Sie jetzt. [Ab nun **3 Minuten**].

Nach 3 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionssseite zurück.“

9. Nun geht es zuerst weiter mit Test 10.

Für diesen Test müssen Sie nun die Spalten von Test 9 und 10 vergleichen und jedes Zeichen, das in der Spalte von Test 10 nicht mit dem entsprechenden Zeichen in Test 9 übereinstimmt, markieren. Im ersten Beispiel passt die 3 nicht, da links eine 0 steht. Im zweiten Beispiel steht anstelle des M links auf der rechten Seite eine 0, die daher markiert werden muss. Es gibt Zeilen, die vollständig identisch sind und Zeilen, die sich in mehr als einem Zeichen unterscheiden. Markieren Sie in diesem Fall beide nicht passenden Zeichen. Versehentlich gemachte Kreuze können durch einen Kreis berichtigt werden. Beachten Sie bitte, dass Sie immer in der rechten Spalte, also in Test 10 die nicht passenden Zeichen markieren.

Fragen?

Für diese Aufgabe haben Sie nur zwei Minuten Zeit. Bitte drehen Sie das Arbeitsblatt nun um und beginnen Sie jetzt. [Ab nun **2 Minuten**].

Nach 2 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionssseite zurück.“

10. Nun kommen wir erst zu Test 9.

In diesem Test ist es ihre Aufgabe, die Zeilen in der Spalte 9 zeilenweise von links nach rechts durchzusehen und zunächst jede achte 0 zu markieren. Beginnen Sie links oben und zählen Sie nur die 0. Jede achte 0 markieren Sie wie im Beispiel mit einem Kreuz. Wenn Sie die gesamte Spalte durchgezählt haben, beginnen Sie wieder von vorne und markieren Sie dann jede achte 1. Beginnen Sie wieder links oben und starten Sie beim Zählen mit eins, unabhängig davon, die wievielte 0 am Ende stand. Wenn sie erneut an das Ende gelangen, zählen Sie jede achte 2, dann jede achte 3 usw. Versehentlich gemachte Kreuze können durch einen Kreis rückgängig gemacht werden. Sollten Sie bis zur 9 kommen, kontrollieren anschließend Ihre Markierungen der 0, 1 usw. Bitte beachten Sie nur die Zahlen und Buchstaben in Spalte 9.

Fragen?

Für diese Aufgabe haben Sie fünf Minuten Zeit. Bitte drehen Sie das Arbeitsblatt nun um und beginnen Sie jetzt, die Nullen in der Spalte und der 9 zu zählen. [Ab nun **5 Minuten**].

Nach 5 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionsseite zurück.“

11. Zum Abschluss zu Test 11.

In diesem Test besteht jede Reihe aus zehn Zahlen, deren Summe sie bilden sollen. Das Ergebnis ist eine zweistellige Zahl. Markieren sie dann die Zahl in der Zeile, die der Einerziffer des Ergebnisses entspricht. Die Summe $2+4+2+6+2+4+2+6+2+5$ ergibt 35. Daher ist die 5 zu markieren. Für das zweite Beispiel gilt: $8+2+6+3+4+2+6+4+2+6=43$. Daher ist die 3 anzukreuzen. Markieren Sie bitte nur die Einerziffer der Lösung. Es werden nur solche Aufgaben gewertet, die genau ein eindeutiges Kreuz haben. Versehentlich gemachte Kreuze können durch einen Kreis berichtigt werden.

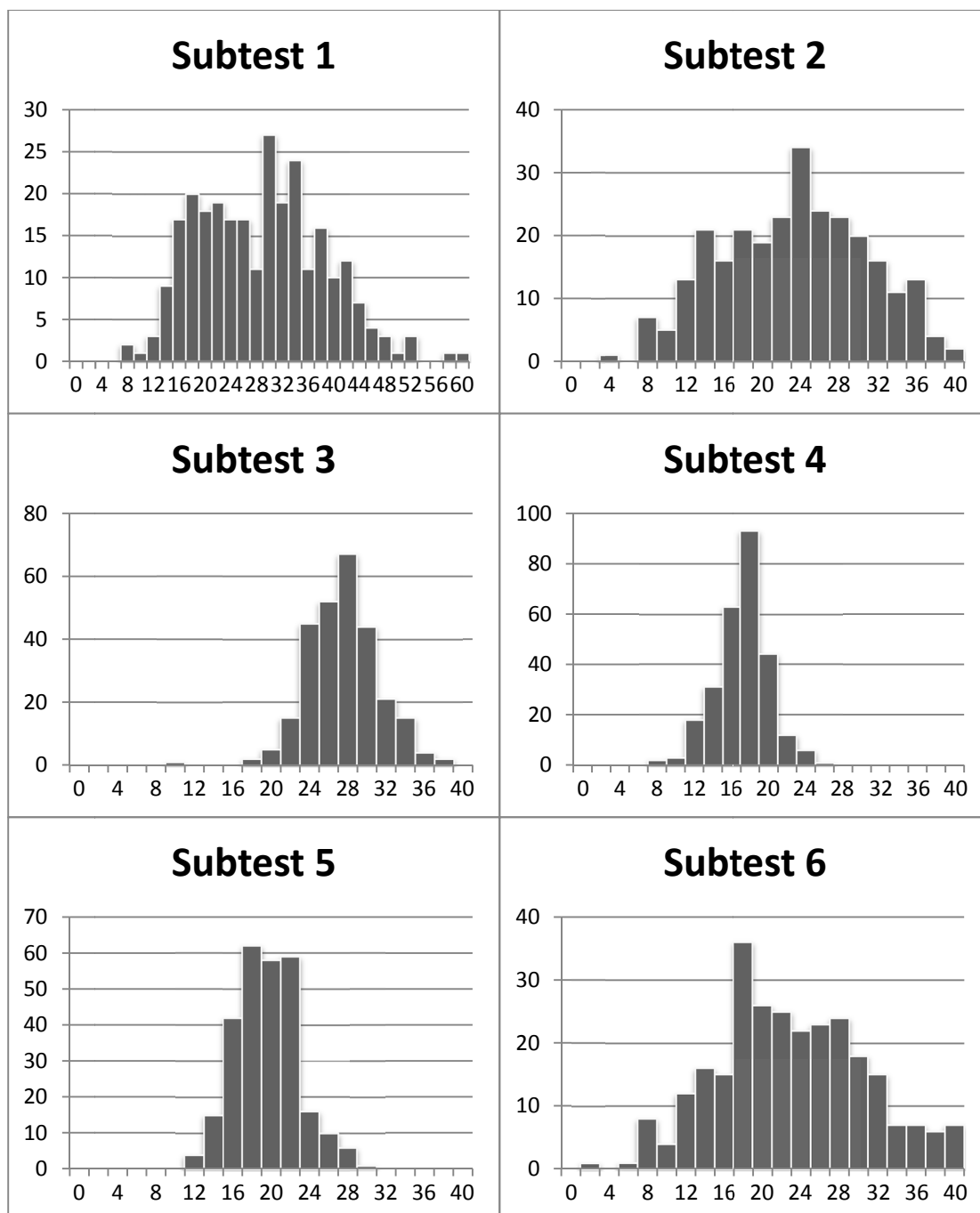
Fragen?

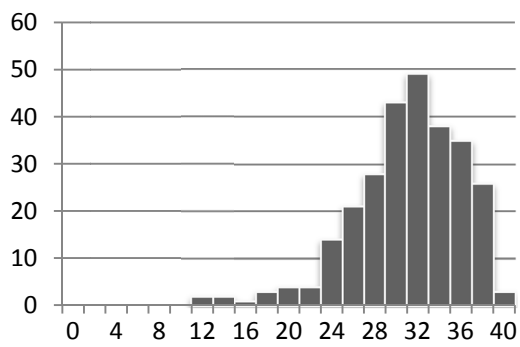
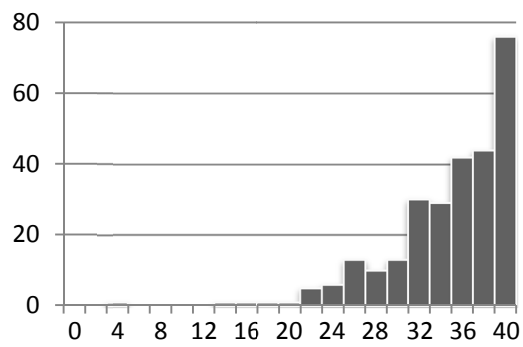
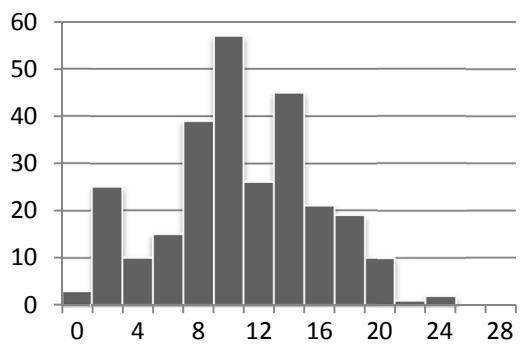
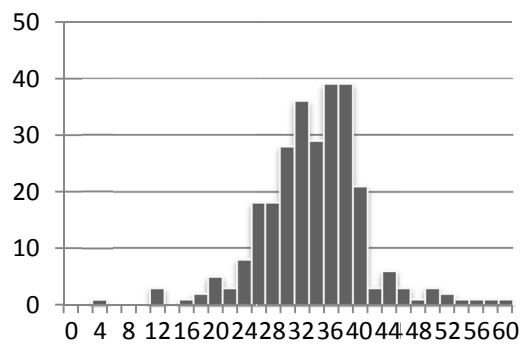
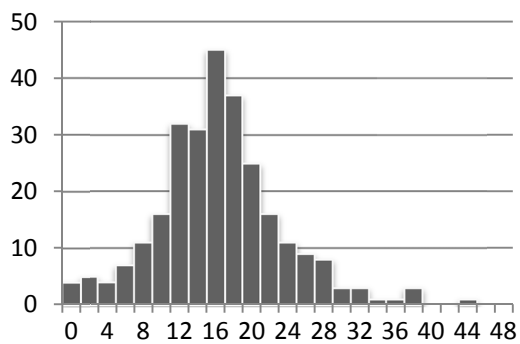
Für diese Aufgabe haben Sie sechs Minuten Zeit. Bitte drehen Sie das Arbeitsblatt nun um und beginnen Sie jetzt. [Ab nun **6 Minuten**].

Nach 6 Minuten: „Halt! Bitte legen Sie den Stift auf die Seite und blättern Sie zur Instruktionsseite zurück.“

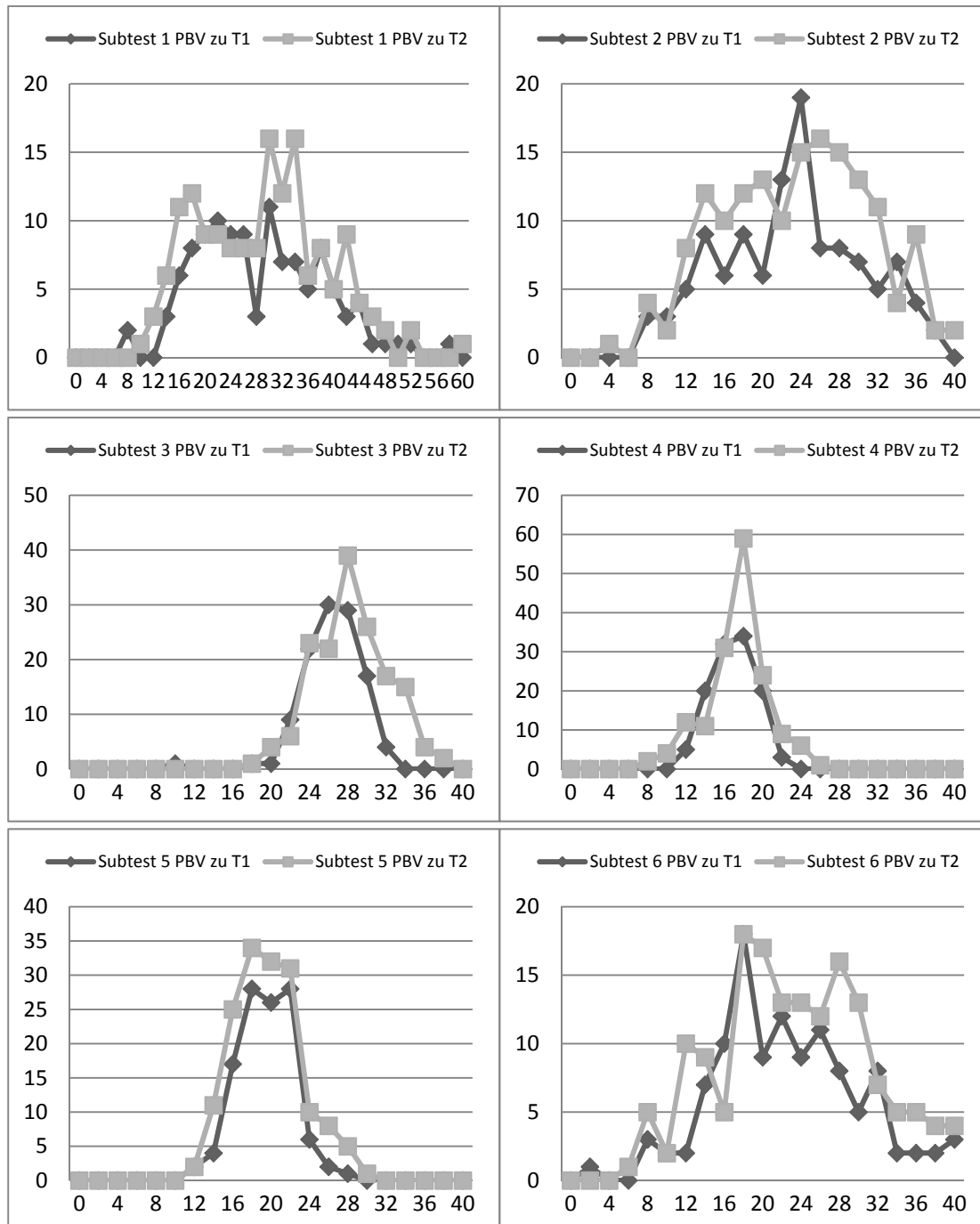
Vielen Dank, dass sie mitgemacht haben.

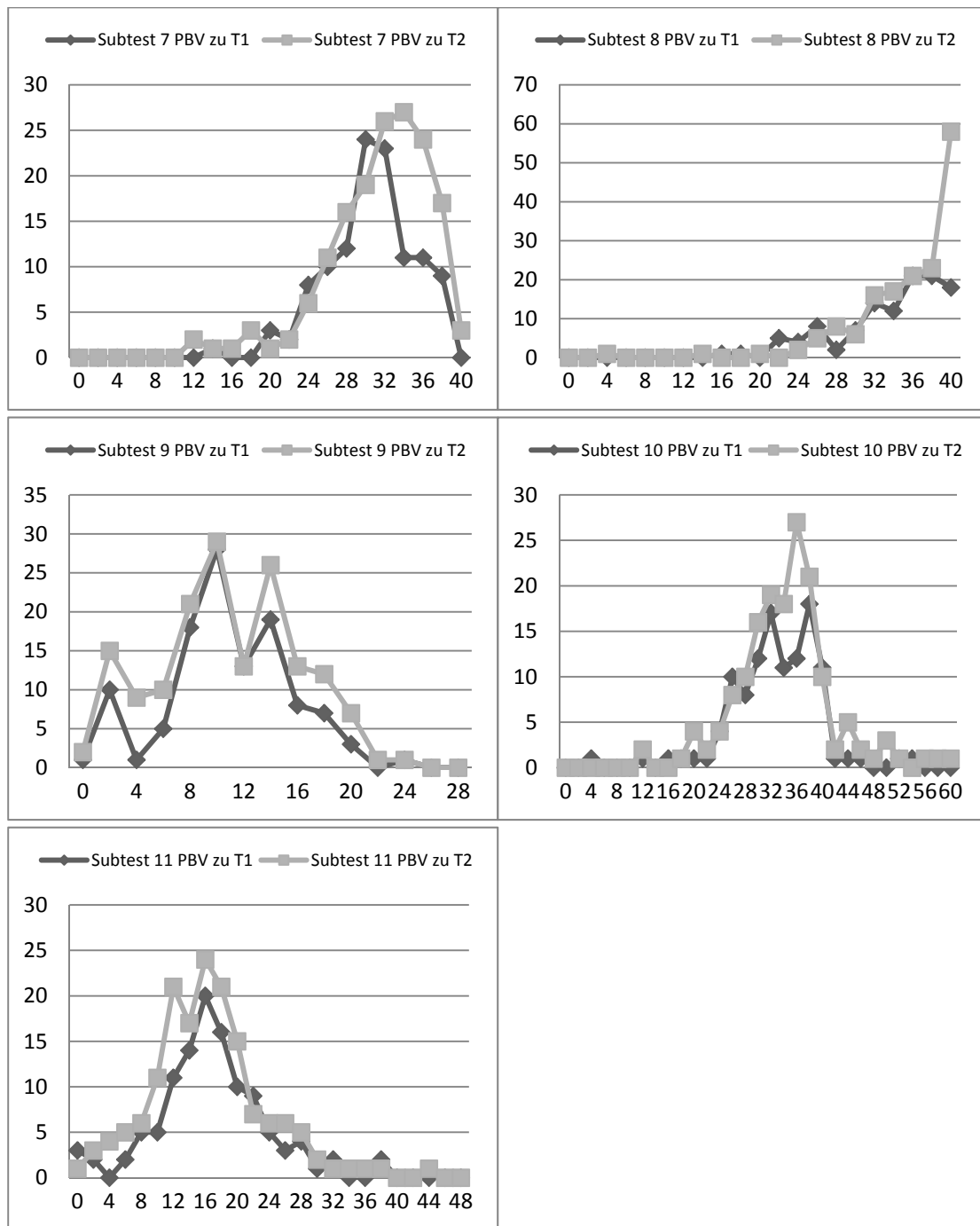
15.3 Verteilungen der Papier-Bleistift Version der Berufsschüler-SP



Subtest 7**Subtest 8****Subtest 9****Subtest 10****Subtest 11**

15.4 Verteilungen der Papier-Bleistift Version der Berufsschüler-SP getrennt nach PBV zu T1 und PBV zu T2





15.5 Itemkennwerte im Kontext der KTT

Für jedes Item wird die Anzahl der Probanden, die es richtig gelöst haben, die es falsch gelöst haben, die das Item ausgelassen haben, aber im Testbogen weitere Items bearbeitet haben und die das Item innerhalb der Zeit nicht mehr bearbeitet haben, erschlossen daraus, dass für kein späteres Item eine Lösung markiert wurde, angegeben. Es werden drei Schwierigkeitsindizes angeführt: p_i gibt das Verhältnis der richtigen Antworten zu allen Versuchspersonen ($n = 273$) an. Hierin spiegelt sich die Ansicht wieder, dass auch Aufgaben, die aufgrund der Zeitbegrenzung nicht mehr bearbeitet werden konnten, als falsch gelöst anzusehen sind. p_i^* stellt das Verhältnis von richtigen Lösungen eines Items zu allen bearbeiteten Items dar. Ein Proband, der ein Item ausgelassen hat, hat dieses Item demnach nicht bearbeitet, weil er die Lösung nicht wusste. p_i^{**} wird durch das Verhältnis von richtig gelösten Items und markierten Items gebildet. Nur wenn eine Entscheidung für eine Lösung getroffen wurde, wird das Item in diesem Fall als bearbeitet angesehen. Insbesondere für p_i^{**} , aber auch für p_i^* hat die Position des Items einen Einfluss auf seine Schwierigkeit. Gulliksen (1950) bezeichnet diese Bestimmung der Schwierigkeit für speeded Tests daher als unangemessen. Da mit Ausnahme der Subtests 4, 5 und 11 es aber einzelnen Probanden durchaus möglich war, in der vorgegeben Zeit den Test zu beenden, wird die Bestimmung der Schwierigkeit durch p_i und die darin begründete Sichtweise als gerechtfertigt angesehen. Als letztes wird für jedes Item die Trennschärfe $r_{it(i)}$, bestimmt nach Formel (8-20), angegeben.

15.5.1 Itemkennwerte Subtest 1

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	270	1	2	0	.99	.99	1.00	.11
2	270	1	2	0	.99	.99	1.00	.16
3	269	2	2	0	.99	.99	.99	.14
4	263	9	1	0	.96	.96	.97	.10
5	256	7	10	0	.94	.94	.97	.04
6	262	8	3	0	.96	.96	.97	.20
7	266	4	3	0	.97	.97	.99	.16
8	240	10	23	0	.88	.88	.96	.02
9	245	16	12	0	.90	.90	.94	.12
10	260	10	3	0	.95	.95	.96	.07
11	253	10	10	0	.93	.93	.96	.04
12	226	28	19	0	.83	.83	.89	.19
13	267	2	4	0	.98	.98	.99	.09
14	142	48	83	0	.52	.52	.75	.18
15	190	12	71	0	.70	.70	.94	.23
16	33	88	152	0	.12	.12	.27	.24
17	204	26	42	1	.75	.75	.89	.40
18	167	22	82	2	.61	.62	.88	.36
19	131	34	102	6	.48	.49	.79	.31
20	117	36	112	8	.43	.44	.76	.39
21	43	58	162	10	.16	.16	.43	.33
22	206	12	44	11	.75	.79	.94	.41
23	103	47	106	17	.38	.40	.69	.14
24	49	67	137	20	.18	.19	.42	.32
25	47	40	159	27	.17	.19	.54	.31
26	11	104	125	33	.04	.05	.10	.25
27	20	43	170	40	.07	.09	.32	.22
28	181	10	46	36	.66	.76	.95	.49
29	56	24	142	51	.21	.25	.70	.24
30	136	9	77	51	.50	.61	.94	.54
31	55	27	130	61	.20	.26	.67	.47
32	81	21	110	61	.30	.38	.79	.54
33	63	93	54	63	.23	.30	.40	.35
34	187	3	16	67	.68	.91	.98	.64
35	186	1	12	74	.68	.93	.99	.68
36	183	2	13	75	.67	.92	.99	.68
37	183	0	11	79	.67	.94	1.00	.68
38	178	2	11	82	.65	.93	.99	.68
39	162	2	20	89	.59	.88	.99	.72
40	158	1	21	93	.58	.88	.99	.68
41	156	3	14	100	.57	.90	.98	.67
42	124	11	24	114	.45	.78	.92	.65
43	106	5	42	120	.39	.69	.95	.63
44	21	56	67	129	.08	.15	.27	.30
45	95	8	38	132	.35	.67	.92	.63
46	50	12	63	148	.18	.40	.81	.57
47	92	6	27	148	.34	.74	.94	.61
48	91	9	20	153	.33	.76	.91	.58
49	26	11	62	174	.10	.26	.70	.48
50	27	8	61	177	.10	.28	.77	.39
51	29	4	58	182	.11	.32	.88	.50
52	22	7	59	185	.08	.25	.76	.46
53	21	4	61	187	.08	.24	.84	.43
54	22	8	54	189	.08	.26	.73	.47
55	11	23	48	191	.04	.13	.32	.29
56	38	13	29	193	.14	.48	.75	.39
57	31	6	40	196	.11	.40	.84	.40
58	55	2	20	196	.20	.71	.96	.43
59	33	7	35	198	.12	.44	.83	.43
60	65	5	0	203	.24	.93	.93	.38

15.5.2 Itemkennwerte Subtest 2

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	267	3	3	0	.98	.98	.99	.02
2	268	3	2	0	.98	.98	.99	.15
3	237	26	10	0	.87	.87	.90	.08
4	189	63	21	0	.69	.69	.75	.31
5	225	29	19	0	.82	.82	.89	.23
6	164	36	73	0	.60	.60	.82	.35
7	222	14	37	0	.81	.81	.94	.27
8	230	10	33	0	.84	.84	.96	.37
9	195	30	48	0	.71	.71	.87	.25
10	216	11	46	0	.79	.79	.95	.33
11	222	21	30	0	.81	.81	.91	.05
12	222	28	22	1	.81	.82	.89	.36
13	208	25	38	2	.76	.77	.89	.29
14	240	9	22	2	.88	.89	.96	.35
15	225	7	37	4	.82	.84	.97	.46
16	215	28	23	7	.79	.81	.88	.18
17	210	10	42	11	.77	.80	.95	.45
18	228	6	26	13	.84	.88	.97	.51
19	200	21	32	20	.73	.79	.90	.32
20	131	38	76	28	.48	.53	.78	.47
21	194	13	32	34	.71	.81	.94	.57
22	167	18	50	38	.61	.71	.90	.63
23	164	19	45	45	.60	.72	.90	.57
24	158	18	47	50	.58	.71	.90	.65
25	139	31	46	57	.51	.64	.82	.53
26	123	22	56	72	.45	.61	.85	.34
27	130	24	34	85	.48	.69	.84	.54
28	117	16	39	101	.43	.68	.88	.51
29	116	7	37	113	.42	.73	.94	.66
30	77	24	53	119	.28	.50	.76	.49
31	112	22	14	125	.41	.76	.84	.60
32	79	16	28	150	.29	.64	.83	.61
33	79	13	18	163	.29	.72	.86	.64
34	65	16	19	173	.24	.65	.80	.61
35	23	32	34	184	.08	.26	.42	.35
36	55	15	10	193	.20	.69	.79	.51
37	40	13	14	206	.15	.60	.75	.43
38	22	14	15	222	.08	.43	.61	.35
39	27	12	6	228	.10	.60	.69	.35
40	19	13	0	241	.07	.59	.59	.29

15.5.3 Itemkennwerte Subtest 3

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	272	1	0	0	1.00	1.00	1.00	.26
2	271	2	0	0	.99	.99	.99	.09
3	273	0	0	0	1.00	1.00	1.00	.00
4	273	0	0	0	1.00	1.00	1.00	.00
5	273	0	0	0	1.00	1.00	1.00	.00
6	264	6	3	0	.97	.97	.98	.10
7	272	0	1	0	1.00	1.00	1.00	-.02
8	272	1	0	0	1.00	1.00	1.00	.26
9	272	1	0	0	1.00	1.00	1.00	.26
10	273	0	0	0	1.00	1.00	1.00	.00
11	272	0	1	0	1.00	1.00	1.00	-.03
12	267	6	0	0	.98	.98	.98	.18
13	258	12	3	0	.95	.95	.96	.14
14	271	2	0	0	.99	.99	.99	-.08
15	266	6	1	0	.97	.97	.98	.21
16	215	46	11	1	.79	.79	.82	.27
17	215	37	20	1	.79	.79	.85	.18
18	261	9	2	1	.96	.96	.97	.19
19	198	69	5	1	.73	.73	.74	.16
20	248	14	10	1	.91	.91	.95	.22
21	255	6	11	1	.93	.94	.98	.27
22	208	52	12	1	.76	.76	.80	.14
23	227	40	5	1	.83	.83	.85	.26
24	218	48	6	1	.80	.80	.82	.12
25	232	23	12	6	.85	.87	.91	.29
26	197	39	18	19	.72	.78	.83	.45
27	135	71	30	37	.49	.57	.66	.29
28	201	12	13	47	.74	.89	.94	.52
29	103	69	23	78	.38	.53	.60	.41
30	83	52	32	106	.30	.50	.61	.43
31	67	53	25	128	.25	.46	.56	.28
32	68	20	31	154	.25	.57	.77	.45
33	43	52	13	165	.16	.40	.45	.42
34	33	41	15	184	.12	.37	.45	.39
35	33	15	19	206	.12	.49	.69	.40
36	24	12	20	217	.09	.43	.67	.39
37	22	13	12	226	.08	.47	.63	.29
38	20	15	6	232	.07	.49	.57	.16
39	8	14	9	242	.03	.26	.36	.18
40	10	19	0	244	.04	.34	.34	.11

15.5.4 Itemkennwerte Subtest 4

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	271	2	0	0	.99	.99	.99	.11
2	273	0	0	0	1	1	1	.00
3	273	0	0	0	1	1	1	.00
4	273	0	0	0	1	1	1	.00
5	272	1	0	0	1	1	1	.12
6	267	6	0	0	.98	.98	.98	.15
7	270	3	0	0	.99	.99	.99	.09
8	220	46	7	0	.81	.81	.83	.27
9	243	27	3	0	.89	.89	.90	.18
10	253	17	3	0	.93	.93	.94	.21
11	192	45	36	0	.70	.70	.81	.31
12	230	21	22	0	.84	.84	.92	.31
13	258	10	5	0	.95	.95	.96	.24
14	216	40	16	1	.79	.79	.84	.29
15	207	43	18	5	.76	.77	.83	.44
16	179	53	35	6	.66	.67	.77	.30
17	61	110	90	12	.22	.23	.36	.13
18	157	47	49	20	.58	.62	.77	.42
19	149	73	22	29	.55	.61	.67	.35
20	38	135	40	60	.14	.18	.22	.24
21	27	83	71	92	.10	.15	.25	.21
22	66	50	53	104	.24	.39	.57	.26
23	54	71	33	115	.20	.34	.43	.20
24	6	77	43	147	.02	.05	.07	.07
25	15	68	26	164	.05	.14	.18	.28
26	18	58	23	174	.07	.18	.24	.25
27	1	44	35	193	.00	.01	.02	.14
28	6	48	19	200	.02	.08	.11	-.01
29	7	33	22	211	.03	.11	.18	-.04
30	6	30	23	214	.02	.10	.17	.12
31	5	28	23	217	.02	.09	.15	.02
32	17	27	8	221	.06	.33	.39	-.07
33	3	27	11	232	.01	.07	.10	.07
34	1	27	10	235	.00	.03	.04	.07
35	3	21	9	240	.01	.09	.13	.00
36	7	16	9	241	.03	.22	.30	-.13
37	3	15	10	245	.01	.11	.17	.05
38	5	23	2	243	.02	.17	.18	-.23
39	0	20	0	253	.00	.00	.00	.00
40	2	16	0	255	.01	.11	.11	.01

15.5.5 Itemkennwerte Subtest 5

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	273	0	0	0	1.00	1.00	1.00	.00
2	271	2	0	0	.99	.99	.99	.03
3	273	0	0	0	1.00	1.00	1.00	.00
4	273	0	0	0	1.00	1.00	1.00	.00
5	272	1	0	0	1.00	1.00	1.00	.04
6	273	0	0	0	1.00	1.00	1.00	.00
7	271	2	0	0	.99	.99	.99	.09
8	270	2	1	0	.99	.99	.99	.12
9	262	10	1	0	.96	.96	.96	.09
10	270	3	0	0	.99	.99	.99	.02
11	270	2	1	0	.99	.99	.99	.09
12	247	25	1	0	.90	.90	.91	.16
13	223	37	12	1	.82	.82	.86	.20
14	234	23	15	1	.86	.86	.91	.34
15	255	14	3	1	.93	.94	.95	.25
16	155	103	13	2	.57	.57	.60	.41
17	102	94	69	8	.37	.38	.52	.40
18	159	66	40	8	.58	.60	.71	.36
19	93	113	51	16	.34	.36	.45	.34
20	144	78	25	26	.53	.58	.65	.38
21	218	19	6	30	.80	.90	.92	.38
22	7	156	57	53	.03	.03	.04	.16
23	93	60	45	75	.34	.47	.61	.39
24	65	104	15	89	.24	.35	.38	.40
25	18	67	52	136	.07	.13	.21	.20
26	53	52	19	149	.19	.43	.50	.36
27	38	56	16	163	.14	.35	.40	.38
28	18	46	22	187	.07	.21	.28	.22
29	15	38	20	200	.05	.21	.28	.27
30	41	18	8	206	.15	.61	.69	.36
31	13	28	9	223	.05	.26	.32	.20
32	3	32	5	233	.01	.08	.09	-.03
33	9	20	6	238	.03	.26	.31	.15
34	5	22	6	240	.02	.15	.19	.23
35	2	23	5	243	.01	.07	.08	.05
36	7	18	3	245	.03	.25	.28	-.01
37	4	21	1	247	.01	.15	.16	-.02
38	0	24	2	247	.00	.00	.00	.00
39	3	18	2	250	.01	.13	.14	-.07
40	2	20	0	251	.01	.09	.09	-.01

15.5.6 Itemkennwerte Subtest 6

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	261	10	2	0	.96	.96	.96	.14
2	269	2	2	0	.99	.99	.99	.12
3	267	4	2	0	.98	.98	.99	.06
4	266	6	1	0	.97	.97	.98	.24
5	269	3	1	0	.99	.99	.99	.00
6	239	33	1	0	.88	.88	.88	.31
7	256	14	3	0	.94	.94	.95	.20
8	235	36	2	0	.86	.86	.87	.17
9	237	32	3	1	.87	.87	.88	.21
10	243	23	6	1	.89	.89	.91	.32
11	253	15	3	2	.93	.93	.94	.34
12	160	97	9	7	.59	.60	.62	.33
13	236	23	4	10	.86	.90	.91	.40
14	234	16	10	13	.86	.90	.94	.41
15	232	21	3	17	.85	.91	.92	.47
16	213	34	2	24	.78	.86	.86	.51
17	218	26	2	27	.80	.89	.89	.51
18	227	11	2	33	.83	.95	.95	.56
19	195	21	3	54	.71	.89	.90	.57
20	172	25	5	71	.63	.85	.87	.56
21	179	12	3	79	.66	.92	.94	.69
22	138	35	5	95	.51	.78	.80	.57
23	140	20	3	110	.51	.86	.88	.69
24	128	11	7	127	.47	.88	.92	.70
25	120	12	5	136	.44	.88	.91	.72
26	106	16	8	143	.39	.82	.87	.71
27	100	12	4	157	.37	.86	.89	.76
28	87	8	5	173	.32	.87	.92	.71
29	74	10	5	184	.27	.83	.88	.66
30	58	12	8	195	.21	.74	.83	.60
31	58	7	4	204	.21	.84	.89	.67
32	34	17	6	216	.12	.60	.67	.53
33	38	11	4	220	.14	.72	.78	.57
34	35	8	7	223	.13	.70	.81	.55
35	34	4	5	230	.12	.79	.89	.53
36	24	10	4	235	.09	.63	.71	.52
37	24	7	3	239	.09	.71	.77	.49
38	23	5	3	242	.08	.74	.82	.47
39	19	6	4	244	.07	.66	.76	.45
40	22	5	0	246	.08	.81	.81	.38

15.5.7 Itemkennwerte Subtest 7

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	267	6	0	0	.98	.98	.98	.11
2	270	3	0	0	.99	.99	.99	.18
3	255	16	2	0	.93	.93	.94	.20
4	271	2	0	0	.99	.99	.99	.01
5	270	2	1	0	.99	.99	.99	.23
6	265	8	0	0	.97	.97	.97	.17
7	265	7	1	0	.97	.97	.97	.18
8	270	3	0	0	.99	.99	.99	.06
9	262	10	1	0	.96	.96	.96	.30
10	263	9	1	0	.96	.96	.97	.12
11	241	30	2	0	.88	.88	.89	.22
12	257	14	2	0	.94	.94	.95	.30
13	265	5	3	0	.97	.97	.98	.21
14	265	4	4	0	.97	.97	.99	.39
15	260	9	4	0	.95	.95	.97	.32
16	242	27	4	0	.89	.89	.90	.32
17	257	9	7	0	.94	.94	.97	.37
18	242	20	10	1	.89	.89	.92	.41
19	251	11	9	2	.92	.93	.96	.38
20	249	11	11	2	.91	.92	.96	.38
21	226	45	0	2	.83	.83	.83	.27
22	249	21	0	3	.91	.92	.92	.32
23	251	18	1	3	.92	.93	.93	.32
24	204	63	1	5	.75	.76	.76	.45
25	236	31	0	6	.86	.88	.88	.46
26	222	42	0	9	.81	.84	.84	.36
27	221	40	0	12	.81	.85	.85	.43
28	237	17	1	18	.87	.93	.93	.49
29	187	67	0	19	.68	.74	.74	.46
30	220	26	2	25	.81	.89	.89	.49
31	104	140	0	29	.38	.43	.43	.35
32	138	96	0	39	.51	.59	.59	.49
33	61	153	3	56	.22	.28	.29	.23
34	134	66	0	73	.49	.67	.67	.47
35	136	42	1	94	.50	.76	.76	.49
36	94	68	2	109	.34	.57	.58	.45
37	71	61	3	138	.26	.53	.54	.50
38	82	37	4	150	.30	.67	.69	.45
39	57	49	3	164	.21	.52	.54	.33
40	29	68	0	176	.11	.30	.30	.25

15.5.8 Itemkennwerte Subtest 8

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	273	0	0	0	1.00	1.00	1.00	.00
2	272	0	1	0	1.00	1.00	1.00	.06
3	270	2	1	0	.99	.99	.99	.03
4	271	1	1	0	.99	.99	1.00	-.05
5	272	0	0	1	1.00	1.00	1.00	.32
6	269	2	1	1	.99	.99	.99	.30
7	271	1	0	1	.99	1.00	1.00	.29
8	270	2	0	1	.99	.99	.99	.30
9	268	4	0	1	.98	.99	.99	.09
10	265	6	1	1	.97	.97	.98	.21
11	269	3	0	1	.99	.99	.99	.30
12	239	25	8	1	.88	.88	.91	.23
13	266	6	0	1	.97	.98	.98	.28
14	268	2	2	1	.98	.99	.99	.32
15	261	11	0	1	.96	.96	.96	.38
16	258	12	1	2	.95	.95	.96	.41
17	256	11	4	2	.94	.94	.96	.43
18	252	16	3	2	.92	.93	.94	.37
19	246	19	6	2	.90	.91	.93	.42
20	258	12	0	3	.95	.96	.96	.46
21	250	19	1	3	.92	.93	.93	.36
22	251	18	1	3	.92	.93	.93	.48
23	260	10	0	3	.95	.96	.96	.48
24	251	17	1	4	.92	.93	.94	.41
25	234	30	2	7	.86	.88	.89	.58
26	222	35	9	7	.81	.83	.86	.58
27	240	22	3	8	.88	.91	.92	.52
28	259	4	2	8	.95	.98	.98	.52
29	210	55	0	8	.77	.79	.79	.48
30	240	15	6	12	.88	.92	.94	.55
31	221	31	5	16	.81	.86	.88	.59
32	224	30	1	18	.82	.88	.88	.66
33	205	34	9	25	.75	.83	.86	.55
34	203	35	3	32	.74	.84	.85	.62
35	156	62	13	42	.57	.68	.72	.57
36	166	53	7	47	.61	.73	.76	.54
37	159	46	11	57	.58	.74	.78	.50
38	106	76	15	76	.39	.54	.58	.48
39	145	39	2	87	.53	.78	.79	.60
40	124	47	0	102	.45	.73	.73	.47

15.5.9 Itemkennwerte Subtest 9

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	257	16	-	-	.94	-	-	.16
2	233	40	-	-	.85	-	-	.24
3	203	70	-	-	.74	-	-	.19
4	179	94	-	-	.66	-	-	.21
5	242	31	-	-	.89	-	-	.31
6	224	49	-	-	.82	-	-	.38
7	204	69	-	-	.75	-	-	.44
8	153	120	-	-	.56	-	-	.36
9	127	146	-	-	.47	-	-	.32
10	94	179	-	-	.34	-	-	.39
11	67	206	-	-	.25	-	-	.40
12	48	225	-	-	.18	-	-	.41
13	203	70	-	-	.74	-	-	.41
14	164	109	-	-	.60	-	-	.50
15	137	136	-	-	.50	-	-	.51
16	107	166	-	-	.39	-	-	.47
17	127	146	-	-	.47	-	-	.40
18	77	196	-	-	.28	-	-	.41
19	49	224	-	-	.18	-	-	.44
20	29	244	-	-	.11	-	-	.32
21	50	223	-	-	.18	-	-	.29
22	29	244	-	-	.11	-	-	.31
23	18	255	-	-	.07	-	-	.34
24	13	260	-	-	.05	-	-	.23
25	12	261	-	-	.04	-	-	.19
26	2	271	-	-	.01	-	-	.12
27	2	271	-	-	.01	-	-	.04
28	2	271	-	-	.01	-	-	.04
29	3	270	-	-	.01	-	-	.01
30	3	270	-	-	.01	-	-	.02
31	0	273	-	-	.00	-	-	.00
32	2	271	-	-	.01	-	-	-.07
33	1	272	-	-	.00	-	-	-.10
34	1	272	-	-	.00	-	-	-.02
35	2	271	-	-	.01	-	-	-.05
36	1	272	-	-	.00	-	-	-.07
37	2	271	-	-	.01	-	-	-.07
38	0	273	-	-	.00	-	-	.00
39	0	273	-	-	.00	-	-	.00
40	0	273	-	-	.00	-	-	.00

15.5.10 Itemkennwerte Subtest 10

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	271	2	-	0	.99	.99	-	.03
2	271	2	-	0	.99	.99	-	.11
3	270	3	-	0	.99	.99	-	.15
4	272	1	-	0	1.00	1.00	-	.25
5	271	2	-	0	.99	.99	-	.15
6	272	1	-	0	1.00	1.00	-	.25
7	266	7	-	0	.97	.97	-	-.06
8	271	2	-	0	.99	.99	-	.22
9	270	2	-	1	.99	.99	-	.12
10	264	8	-	1	.97	.97	-	.10
11	269	3	-	1	.99	.99	-	.25
12	269	3	-	1	.99	.99	-	.19
13	268	1	-	4	.98	1.00	-	.27
14	267	2	-	4	.98	.99	-	.40
15	218	51	-	4	.80	.81	-	.02
16	269	0	-	4	.99	1.00	-	.39
17	267	2	-	4	.98	.99	-	.33
18	266	3	-	4	.97	.99	-	.34
19	262	6	-	5	.96	.98	-	.37
20	263	2	-	8	.96	.99	-	.46
21	262	0	-	11	.96	1.00	-	.49
22	250	12	-	11	.92	.95	-	.27
23	258	4	-	11	.95	.98	-	.44
24	252	2	-	19	.92	.99	-	.51
25	242	12	-	19	.89	.95	-	.39
26	234	11	-	28	.86	.96	-	.49
27	236	2	-	35	.86	.99	-	.57
28	236	1	-	36	.86	1.00	-	.60
29	217	2	-	54	.79	.99	-	.60
30	202	15	-	56	.74	.93	-	.53
31	186	12	-	75	.68	.94	-	.59
32	144	36	-	93	.53	.80	-	.50
33	158	2	-	113	.58	.99	-	.69
34	132	2	-	139	.48	.99	-	.69
35	131	2	-	140	.48	.98	-	.68
36	113	3	-	157	.41	.97	-	.64
37	72	28	-	173	.26	.72	-	.49
38	73	4	-	196	.27	.95	-	.61
39	63	1	-	209	.23	.98	-	.59
40	37	21	-	215	.14	.64	-	.49
41	26	1	-	246	.10	.96	-	.55
42	26	1	-	246	.10	.96	-	.55
43	26	1	-	246	.10	.96	-	.55
44	26	1	-	246	.10	.96	-	.54
45	15	4	-	254	.05	.79	-	.49
46	12	0	-	261	.04	1.00	-	.51
47	12	0	-	261	.04	1.00	-	.51
48	9	3	-	261	.03	.75	-	.47
49	10	2	-	261	.04	.83	-	.47
50	9	0	-	264	.03	1.00	-	.47
51	9	0	-	264	.03	1.00	-	.47
52	6	3	-	264	.02	.67	-	.40
53	7	1	-	265	.03	.88	-	.43
54	6	1	-	266	.02	.86	-	.41
55	6	0	-	267	.02	1.00	-	.41
56	6	0	-	267	.02	1.00	-	.41
57	6	0	-	267	.02	1.00	-	.41
58	6	0	-	267	.02	1.00	-	.41
59	6	0	-	267	.02	1.00	-	.41
60	6	0	-	267	.02	1.00	-	.41

15.5.11 Itemkennwerte Subtest 11

Item	richtig	falsch	ausgelassen	nicht bearbeitet	p_i	p_i^*	p_i^{**}	$r_{it(i)}$
1	233	38	2	0	.85	.85	.86	.39
2	219	53	1	0	.80	.80	.81	.18
3	226	43	4	0	.83	.83	.84	.36
4	190	81	2	0	.70	.70	.70	.35
5	219	52	2	0	.80	.80	.81	.32
6	243	28	1	1	.89	.89	.90	.44
7	214	56	2	1	.78	.79	.79	.29
8	232	37	2	2	.85	.86	.86	.39
9	233	34	2	4	.85	.87	.87	.42
10	226	40	1	6	.83	.85	.85	.43
11	207	54	3	9	.76	.78	.79	.37
12	214	40	1	18	.78	.84	.84	.36
13	212	32	2	27	.78	.86	.87	.54
14	192	39	1	41	.70	.83	.83	.50
15	180	37	2	54	.66	.82	.83	.51
16	156	41	6	70	.57	.77	.79	.51
17	140	41	1	91	.51	.77	.77	.53
18	131	28	3	111	.48	.81	.82	.61
19	113	25	4	131	.41	.80	.82	.58
20	89	26	5	153	.33	.74	.77	.62
21	77	25	4	167	.28	.73	.75	.59
22	69	17	4	183	.25	.77	.80	.68
23	52	20	3	198	.19	.69	.72	.61
24	46	15	3	209	.17	.72	.75	.62
25	39	14	2	218	.14	.71	.74	.61
26	31	16	3	223	.11	.62	.66	.59
27	26	13	3	231	.10	.62	.67	.60
28	22	11	3	237	.08	.61	.67	.54
29	20	10	3	240	.07	.61	.67	.55
30	15	9	3	246	.05	.56	.63	.44
31	11	10	2	250	.04	.48	.52	.42
32	10	9	2	252	.04	.48	.53	.44
33	11	7	1	254	.04	.58	.61	.37
34	8	9	2	254	.03	.42	.47	.39
35	7	7	2	257	.03	.44	.50	.38
36	7	6	2	258	.03	.47	.54	.39
37	7	6	2	258	.03	.47	.54	.36
38	6	7	1	259	.02	.43	.46	.44
39	4	6	1	262	.01	.36	.40	.38
40	3	7	0	263	.01	.30	.30	.34
41	2	5	1	265	.01	.25	.29	.29
42	1	5	1	266	.00	.14	.17	.24
43	1	5	1	266	.00	.14	.17	.24
44	3	3	1	266	.01	.43	.50	.08
45	0	6	0	267	.00	.00	.00	.00

Ab Item 45 wurden zwar einzelne der 80 Items noch angekreuzt, es ist jedoch offensichtlich, dass hierbei nur mehr geraten wurde. Daher werden für die restlichen 35 Items keine Kennwerte angegeben.

15.6 ANOVA Tabellen der Computerbedingungen

In den folgenden Tabellen werden die Ergebnisse der univariaten Varianzanalysen der Ergebnisse der elf Subtests des LPS-neu für die vier Faktoren Präsentation (einzeln, alle), Antwortformat (einmalig, weiter, korrigierbar), Zeitvorgabe (keine, gesamt, einzeln) und Zeitpunkt (CV zu T1 vs. CV zu T2) dargestellt. Angegeben werden die Quadratsummen (QS), die Freiheitsgrade (df ; df_2 entspricht df_{gesamt}), die mittleren Quadratsummen ($MQS = QS/df$), die entsprechenden F -Werte und p -Values, das klassische Varianzverhältnis (η_c^2 ; Varianzklärung im Kontext des Gesamtmodells), das partielle Varianzverhältnis (η_p^2 ; Varianzklärung ohne Berücksichtigung der restlichen Faktoren) und die untere und obere Grenze des 80 %-Konfidenzintervalls von η_p^2 .

15.6.1 Subtest 1

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	338.14	1	338.14	4.46	.04	.012	.018	.003	.046
Antwortformat (einmalig, weiter, korr.)	282.44	2	141.22	1.86	.16	.010	.015	0	.036
Zeitvorgabe (keine, gesamt, einzeln)	73.72	2	36.86	0.49	.62	.003	.004	0	.014
Zeitpunkt (CV zu T1 vs. CV zu T2)	6112.86	1	6112.86	80.55	<.01	.215	.253	.193	.309
Präsentation * Antwortformat	170.05	2	85.02	1.12	.33	.006	.009	0	.025
Präsentation * Zeitvorgabe	48.05	1	48.05	0.63	.43	.002	.003	0	.017
Präsentation * Zeitpunkt	77.90	1	77.90	1.03	.31	.003	.004	0	.021
Antwortformat * Zeitvorgabe	486.85	4	121.71	1.60	.17	.017	.026	0	.043
Antwortformat * Zeitpunkt	159.19	2	79.59	1.05	.35	.006	.009	0	.024
Zeitvorgabe * Zeitpunkt	922.64	2	461.32	6.08	<.01	.032	.049	.016	.083
Präs. * Antwortformat * Zeitvorgabe	414.36	2	207.18	2.73	.07	.015	.022	.001	.047
Präsentation * Antwortformat * Zeitpunkt	277.15	2	138.57	1.83	.16	.010	.015	0	.036
Präsentation * Zeitvorgabe * Zeitpunkt	17.14	1	17.14	0.23	.64	<.001	.001	0	.012
Antwortformat * Zeitvorgabe * Zeitpunkt	494.40	4	123.60	1.63	.17	.017	.027	0	.043
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	534.02	2	267.01	3.52	.03	.019	.029	.005	.056
Residual	18061.09	238	75.89			.634			
Gesamt	28470.00	268							

15.6.2 Subtest 2

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	10.84	1	10.84	0.21	.64	<.001	.001	0	.011
Antwortformat (einmalig, weiter, korr.)	33.93	2	16.96	0.34	.72	.002	.003	0	.010
Zeitvorgabe (keine, gesamt, einzeln)	390.27	2	195.14	3.86	.02	.022	.031	.006	.060
Zeitpunkt (CV zu T1 vs. CV zu T2)	4110.40	1	4110.40	81.35	<.01	.233	.254	.194	.310
Präsentation * Antwortformat	89.26	2	44.63	0.88	.42	.005	.007	0	.022
Präsentation * Zeitvorgabe	152.50	1	152.50	3.02	.08	.009	.012	0	.037
Präsentation * Zeitpunkt	45.66	1	45.66	0.90	.34	.003	.004	0	.020
Antwortformat * Zeitvorgabe	126.78	4	31.69	0.63	.64	.007	.010	0	.016
Antwortformat * Zeitpunkt	4.99	2	2.49	0.05	.95	<.001	<.001	0	0
Zeitvorgabe * Zeitpunkt	110.58	2	55.29	1.09	.34	.006	.009	0	.025
Präs. * Antwortformat * Zeitvorgabe	166.35	2	83.17	1.65	.20	.009	.014	0	.033
Präsentation * Antwortformat * Zeitpunkt	57.03	2	28.52	0.56	.57	.003	.005	0	.016
Präsentation * Zeitvorgabe * Zeitpunkt	29.77	1	29.77	0.59	.44	.002	.002	0	.017
Antwortformat * Zeitvorgabe * Zeitpunkt	143.01	4	35.75	0.71	.59	.008	.012	0	.019
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	87.50	2	43.75	0.87	.42	.005	.007	0	.021
Residual	12076.62	239	50.53			.685			
Gesamt	17635.48	269							

15.6.3 Subtest 3

Faktor	<i>QS</i>	<i>df</i>	<i>MQS</i>	<i>F</i>	<i>p</i>	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	14.09	1	14.09	1.06	.30	.003	.004	0	.022
Antwortformat (einmalig, weiter, korr.)	8.39	2	4.19	0.32	.73	.002	.003	0	.010
Zeitvorgabe (keine, gesamt, einzeln)	67.26	2	33.63	2.54	.08	.017	.021	0	.045
Zeitpunkt (CV zu T1 vs. CV zu T2)	456.65	1	456.65	34.48	<.01	.113	.127	.079	.177
Präsentation * Antwortformat	36.82	2	18.41	1.39	.25	.009	.012	0	.030
Präsentation * Zeitvorgabe	23.33	1	23.33	1.76	.19	.006	.007	0	.028
Präsentation * Zeitpunkt	0.59	1	0.59	0.05	.83	<.001	<.001	0	.005
Antwortformat * Zeitvorgabe	30.40	4	7.60	0.57	.68	.008	.010	0	.014
Antwortformat * Zeitpunkt	68.57	2	34.28	2.59	.08	.017	.021	0	.045
Zeitvorgabe * Zeitpunkt	7.63	2	3.81	0.29	.75	.002	.002	0	.009
Präs. * Antwortformat * Zeitvorgabe	11.84	2	5.92	0.45	.64	.003	.004	0	.013
Präsentation * Antwortformat * Zeitpunkt	88.11	2	44.05	3.33	.04	.022	.027	.004	.054
Präsentation * Zeitvorgabe * Zeitpunkt	23.18	1	23.18	1.75	.19	.006	.007	0	.028
Antwortformat * Zeitvorgabe * Zeitpunkt	39.63	4	9.91	0.75	.56	.010	.012	0	.020
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	3.25	2	1.62	0.12	.89	<.001	.001	0	.001
Residual	3151.88	238	13.24			.782			
Gesamt	4031.60	268							

15.6.4 Subtest 4

Faktor	<i>QS</i>	<i>df</i>	<i>MQS</i>	<i>F</i>	<i>p</i>	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	0.18	1	0.18	0.02	.89	<.001	<.001	0	0
Antwortformat (einmalig, weiter, korr.)	22.90	2	11.45	1.30	.28	.009	.011	0	.028
Zeitvorgabe (keine, gesamt, einzeln)	35.84	2	17.92	2.03	.13	.014	.017	0	.038
Zeitpunkt (CV zu T1 vs. CV zu T2)	103.71	1	103.71	11.77	<.01	.040	.047	.018	.084
Präsentation * Antwortformat	26.75	2	13.37	1.52	.22	.010	.012	0	.031
Präsentation * Zeitvorgabe	0.26	1	0.26	0.03	.87	<.001	<.001	0	.003
Präsentation * Zeitpunkt	13.74	1	13.74	1.56	.21	.005	.006	0	.026
Antwortformat * Zeitvorgabe	22.75	4	5.69	0.65	.63	.009	.011	0	.016
Antwortformat * Zeitpunkt	25.61	2	12.80	1.45	.24	.010	.012	0	.030
Zeitvorgabe * Zeitpunkt	87.01	2	43.51	4.94	<.01	.034	.040	.011	.071
Präs. * Antwortformat * Zeitvorgabe	32.09	2	16.04	1.82	.16	.013	.015	0	.035
Präsentation * Antwortformat * Zeitpunkt	13.55	2	6.77	0.77	.47	.005	.006	0	.020
Präsentation * Zeitvorgabe * Zeitpunkt	0.30	1	0.30	0.03	.85	<.001	<.001	0	.003
Antwortformat * Zeitvorgabe * Zeitpunkt	31.23	4	7.81	0.89	.47	.012	.015	0	.024
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	35.85	2	17.92	2.03	.13	.014	.017	0	.038
Residual	2114.55	240	8.81			.824			
Gesamt	2566.29	270							

15.6.5 Subtest 5

Faktor	<i>QS</i>	<i>df</i>	<i>MQS</i>	<i>F</i>	<i>p</i>	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	57.69	1	57.69	5.43	.02	.018	.022	.004	.052
Antwortformat (einmalig, weiter, korr.)	72.61	2	36.31	3.42	.03	.022	.028	.004	.055
Zeitvorgabe (keine, gesamt, einzeln)	79.76	2	39.88	3.75	.03	.024	.031	.005	.059
Zeitpunkt (CV zu T1 vs. CV zu T2)	65.25	1	65.25	6.14	.01	.020	.025	.006	.056
Präsentation * Antwortformat	18.68	2	9.34	0.88	.42	.006	.007	0	.022
Präsentation * Zeitvorgabe	26.10	1	26.10	2.46	.12	.008	.010	0	.033
Präsentation * Zeitpunkt	0.04	1	0.04	0.00	.95	<.001	<.001	0	0
Antwortformat * Zeitvorgabe	100.62	4	25.16	2.37	.05	.031	.038	.003	.060
Antwortformat * Zeitpunkt	38.55	2	19.28	1.81	.17	.012	.015	0	.035
Zeitvorgabe * Zeitpunkt	13.94	2	6.97	0.66	.52	.004	.005	0	.018
Präs. * Antwortformat * Zeitvorgabe	13.62	2	6.81	0.64	.53	.004	.005	0	.017
Präsentation * Antwortformat * Zeitpunkt	59.85	2	29.92	2.82	.06	.018	.023	.002	.048
Präsentation * Zeitvorgabe * Zeitpunkt	0.07	1	0.07	0.01	.94	<.001	<.001	0	.002
Antwortformat * Zeitvorgabe * Zeitpunkt	194.91	4	48.73	4.58	<.01	.059	.072	.025	.102
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	19.95	2	9.97	0.94	.39	.006	.008	0	.023
Residual	2529.79	238	10.63			.769			
Gesamt	3291.43	268							

15.6.6 Subtest 6

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	8.09	1	8.09	0.19	.66	<.001	.001	0	.011
Antwortformat (einmalig, weiter, korr.)	16.64	2	8.32	0.20	.82	.001	.002	0	.005
Zeitvorgabe (keine, gesamt, einzeln)	131.81	2	65.90	1.54	.22	.010	.013	0	.032
Zeitpunkt (CV zu T1 vs. CV zu T2)	1493.52	1	1493.52	34.96	<.01	.117	.128	.080	.179
Präsentation * Antwortformat	90.49	2	45.25	1.06	.35	.007	.009	0	.025
Präsentation * Zeitvorgabe	59.30	1	59.30	1.39	.24	.005	.006	0	.025
Präsentation * Zeitpunkt	12.83	1	12.83	0.30	.58	.001	.001	0	.013
Antwortformat * Zeitvorgabe	306.02	4	76.50	1.79	.13	.024	.029	0	.047
Antwortformat * Zeitpunkt	6.95	2	3.48	0.08	.92	<.001	.001	0	.002
Zeitvorgabe * Zeitpunkt	59.36	2	29.68	0.70	.50	.005	.006	0	.018
Präs. * Antwortformat * Zeitvorgabe	11.84	2	5.92	0.14	.87	<.001	.001	0	.002
Präsentation * Antwortformat * Zeitpunkt	90.75	2	45.38	1.06	.35	.007	.009	0	.025
Präsentation * Zeitvorgabe * Zeitpunkt	0.71	1	0.71	0.02	.90	<.001	<.001	0	0
Antwortformat * Zeitvorgabe * Zeitpunkt	227.84	4	56.96	1.33	.26	.018	.022	0	.036
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	33.59	2	16.79	0.39	.68	.003	.003	0	.012
Residual	10169.05	238	42.73			.800			
Gesamt	12718.78	268							

15.6.7 Subtest 7

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	4.54	1	4.54	0.22	.64	<.001	.001	0	.012
Antwortformat (einmalig, weiter, korr.)	1.41	2	0.71	0.04	.97	<.001	<.001	0	0
Zeitvorgabe (keine, gesamt, einzeln)	180.42	2	90.21	4.44	.01	.032	.037	.009	.068
Zeitpunkt (CV zu T1 vs. CV zu T2)	423.75	1	423.75	20.83	<.01	.074	.083	.044	.129
Präsentation * Antwortformat	5.35	2	2.67	0.13	.88	<.001	.001	0	.002
Präsentation * Zeitvorgabe	47.50	1	47.50	2.34	.13	.008	.010	0	.033
Präsentation * Zeitpunkt	117.99	1	117.99	5.80	.02	.021	.025	.005	.056
Antwortformat * Zeitvorgabe	21.65	4	5.41	0.27	.90	.004	.005	0	.002
Antwortformat * Zeitpunkt	62.45	2	31.23	1.54	.22	.011	.013	0	.033
Zeitvorgabe * Zeitpunkt	2.56	2	1.28	0.06	.94	<.001	.001	0	.001
Präs. * Antwortformat * Zeitvorgabe	27.55	2	13.77	0.68	.51	.005	.006	0	.019
Präsentation * Antwortformat * Zeitpunkt	61.15	2	30.57	1.50	.23	.011	.013	0	.032
Präsentation * Zeitvorgabe * Zeitpunkt	4.49	1	4.49	0.22	.64	<.001	.001	0	.012
Antwortformat * Zeitvorgabe * Zeitpunkt	54.11	4	13.53	0.67	.62	.010	.011	0	.018
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	1.96	2	0.98	0.05	.95	<.001	<.001	0	0
Residual	4678.09	230	20.34			.821			
Gesamt	5694.94	260							

15.6.8 Subtest 8

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	71.72	1	71.72	2.15	.14	.008	.009	0	.031
Antwortformat (einmalig, weiter, korr.)	36.03	2	18.01	0.54	.58	.004	.005	0	.015
Zeitvorgabe (keine, gesamt, einzeln)	101.36	2	50.68	1.52	.22	.011	.013	0	.032
Zeitpunkt (CV zu T1 vs. CV zu T2)	819.78	1	819.78	24.62	<.01	.089	.095	.053	.142
Präsentation * Antwortformat	61.96	2	30.98	0.93	.40	.007	.008	0	.023
Präsentation * Zeitvorgabe	5.58	1	5.58	0.17	.68	<.001	.001	0	.010
Präsentation * Zeitpunkt	7.31	1	7.31	0.22	.64	<.001	.001	0	.012
Antwortformat * Zeitvorgabe	79.51	4	19.88	0.60	.67	.009	.010	0	.015
Antwortformat * Zeitpunkt	4.35	2	2.18	0.07	.94	<.001	.001	0	.002
Zeitvorgabe * Zeitpunkt	4.42	2	2.21	0.07	.94	<.001	.001	0	.002
Präs. * Antwortformat * Zeitvorgabe	21.92	2	10.96	0.33	.72	.002	.003	0	.010
Präsentation * Antwortformat * Zeitpunkt	11.63	2	5.82	0.18	.84	.001	.001	0	.005
Präsentation * Zeitvorgabe * Zeitpunkt	17.84	1	17.84	0.54	.47	.002	.002	0	.017
Antwortformat * Zeitvorgabe * Zeitpunkt	46.37	4	11.59	0.35	.85	.005	.006	0	.006
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	91.45	2	45.72	1.37	.26	.010	.012	0	.030
Residual	7825.39	235	33.30			.850			
Gesamt	9206.59	265							

15.6.9 Subtest 9

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	999.68	1	999.68	65.24	<.01	.163	.217	.159	.273
Antwortformat (einmalig, weiter, korr.)	418.14	2	209.07	13.64	<.01	.068	.104	.057	.149
Zeitvorgabe (keine, gesamt, einzeln)	47.05	1	47.05	3.07	.08	.008	.013	0	.037
Zeitpunkt (CV zu T1 vs. CV zu T2)	199.75	1	199.75	13.04	<.01	.033	.053	.022	.092
Präsentation * Antwortformat	340.08	1	340.08	22.19	<.01	.055	.086	.046	.133
Präsentation * Zeitvorgabe	3.48	1	3.48	0.23	.63	<.001	.001	0	.012
Präsentation * Zeitpunkt	239.15	1	239.15	15.61	<.01	.039	.062	.029	.104
Antwortformat * Zeitvorgabe	146.42	2	73.21	4.78	<.01	.024	.039	.010	.070
Antwortformat * Zeitpunkt	64.49	2	32.25	2.10	.12	.011	.018	0	.040
Zeitvorgabe * Zeitpunkt	6.43	1	6.43	0.42	.52	.001	.002	0	.015
Präs. * Antwortformat * Zeitvorgabe	0.12	1	0.12	0.01	.93	<.001	<.001	0	.002
Präsentation * Antwortformat * Zeitpunkt	17.55	1	17.55	1.15	.29	.003	.005	0	.023
Präsentation * Zeitvorgabe * Zeitpunkt	0.07	1	0.07	0.00	.95	<.001	<.001	0	0
Antwortformat * Zeitvorgabe * Zeitpunkt	35.29	2	17.65	1.15	.32	.006	.010	0	.026
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	22.30	1	22.30	1.46	.23	.004	.006	0	.025
Residual	3600.97	235	15.32			.586			
Gesamt	6140.96	255							

15.6.10 Subtest 10

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	4699.84	1	4699.84	95.30	<.01	.251	.291	.229	.347
Antwortformat (einmalig, weiter, korr.)	422.44	2	211.22	4.28	.02	.023	.036	.008	.066
Zeitvorgabe (keine, gesamt, einzeln)	294.64	1	294.64	5.97	.02	.016	.025	.006	.056
Zeitpunkt (CV zu T1 vs. CV zu T2)	154.48	1	154.48	3.13	.08	.008	.013	0	.038
Präsentation * Antwortformat	13.28	1	13.28	0.27	.60	<.001	.001	0	.013
Präsentation * Zeitvorgabe	7.70	1	7.70	0.16	.69	<.001	.001	0	.010
Präsentation * Zeitpunkt	29.00	1	29.00	0.59	.44	.002	.003	0	.017
Antwortformat * Zeitvorgabe	1195.52	2	597.76	12.12	<.01	.064	.095	.049	.138
Antwortformat * Zeitpunkt	149.73	2	74.86	1.52	.22	.008	.013	0	.032
Zeitvorgabe * Zeitpunkt	68.75	1	68.75	1.39	.24	.004	.006	0	.025
Präs. * Antwortformat * Zeitvorgabe	86.66	1	86.66	1.76	.19	.005	.008	0	.028
Präsentation * Antwortformat * Zeitpunkt	105.08	1	105.08	2.13	.15	.006	.009	0	.031
Präsentation * Zeitvorgabe * Zeitpunkt	4.22	1	4.22	0.09	.77	<.001	<.001	0	.007
Antwortformat * Zeitvorgabe * Zeitpunkt	63.30	2	31.65	0.64	.53	.003	.006	0	.018
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	1.22	1	1.22	0.03	.88	<.001	<.001	0	.003
Residual	11441.55	232	49.32			.611			
Gesamt	18737.40	252							

15.6.11 Subtest 11

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Präsentation (einzeln, alle)	202.08	1	202.08	4.10	.04	.015	.019	.002	.049
Antwortformat (einmalig, weiter, korr.)	68.74	2	34.37	0.70	.50	.005	.007	0	.021
Zeitvorgabe (keine, gesamt, einzeln)	427.79	2	213.89	4.33	.01	.033	.039	.009	.072
Zeitpunkt (CV zu T1 vs. CV zu T2)	1086.94	1	1086.94	22.03	<.01	.083	.094	.050	.144
Präsentation * Antwortformat	14.98	2	7.49	0.15	.86	.001	.001	0	.003
Präsentation * Zeitvorgabe	5.62	1	5.62	0.11	.74	<.001	.001	0	.009
Präsentation * Zeitpunkt	62.38	1	62.38	1.26	.26	.005	.006	0	.026
Antwortformat * Zeitvorgabe	135.95	4	33.99	0.69	.60	.010	.013	0	.020
Antwortformat * Zeitpunkt	52.69	2	26.35	0.53	.59	.004	.005	0	.017
Zeitvorgabe * Zeitpunkt	177.02	2	88.51	1.79	.17	.014	.017	0	.039
Präs. * Antwortformat * Zeitvorgabe	153.31	2	76.65	1.55	.21	.012	.014	0	.036
Präsentation * Antwortformat * Zeitpunkt	44.27	2	22.14	0.45	.64	.003	.004	0	.015
Präsentation * Zeitvorgabe * Zeitpunkt	72.51	1	72.51	1.47	.23	.006	.007	0	.028
Antwortformat * Zeitvorgabe * Zeitpunkt	60.05	4	15.01	0.30	.88	.005	.006	0	.004
Präs. * Antwortf. * Zeitvorgabe * Zeitpunkt	67.20	2	33.60	0.68	.51	.005	.006	0	.020
Residual	10462.30	212	49.35			.799			
Gesamt	13093.81	242							

15.7 Mittelwerte der Rohwertsummen

Subtest 1

Teststufenfolge	Präsentation alle												Gesamt																															
	Antwortformat						Antwortformat						Antwortformat																															
	einmalig		korrigierbar		Gesamt		einmalig		korrigierbar		Gesamt		einmalig		korrigierbar		Gesamt																											
	ohne Gesamt	einzelne	Zetvorgabe	ohne Gesamt	einzelne	Zetvorgabe	ohne Gesamt	einzelne	Zetvorgabe	ohne Gesamt	einzelne	Zetvorgabe	ohne Gesamt	einzelne	Zetvorgabe	ohne Gesamt	einzelne	Zetvorgabe																										
PBW-SD	M 25.1	25.7	31.0	27.2	27.3	24.3	26.1	26.0	30.2	27.3	26.9	28.4	27.2	29.8	26.9	33.0	33.4	30.3	28.9	29.6	31.2	29.8	30.5	27.3	26.4	31.0	27.7	30.6	29.3	26.1	29.3	30.2	28.2	26.9	28.9	29.3	28.0	28.2	28.6					
	n 11.8	7	9	27	9	8	8	25	11	9	7	27	31	24	79	10	100	107	102	111	143	127	114	73	93	106	112	90	98	115	97	101	120	83	106	100	104	70	97	96	107	92	99	
	PC-PB	M 15.6	18.1	24.9	19.4	19.3	18.5	17.8	17.5	16.9	20.8	15.4	17.8	16.3	19.3	18.2	18.3	20.7	19.4	24.3	27.2	25.9	19.7	20.4	20.1	20.7	23.0	21.8	16.9	19.6	24.9	20.3	23.5	17.8	21.2	18.2	20.6	1.5	188	184	21.3	198	198	
PBW-SD	M 37.0	22.1	29.1	29.1	21.8	28.8	29.1	26.6	29.9	27.7	25.9	27.8	28.7	28.2	28.5	31.6	22.4	27.0	32.7	27.7	30.4	30.7	26.3	28.7	32.3	24.7	29.1	28.8	25.5	26.3	29.1	26.7	31.2	27.7	25.9	28.8	29.9	26.2	28.1	28.1				
	n 7	8	8	23	8	8	9	25	8	6	8	22	23	22	25	7	6	15	5	10	7	6	13	21	17	38	16	14	8	38	13	13	9	35	15	12	8	35	44	39	25	108		
	PB-PC	M 43.0	25.6	30.6	32.7	27.0	32.3	29.6	29.6	27.4	30.2	25.6	27.9	34.4	30.3	32.8	34.4	25.4	29.9	35.9	29.5	32.9	34.9	28.6	32.1	38.2	27.6	30.6	32.7	29.8	29.6	29.6	29.7	31.3	29.8	25.6	29.5	33.4	29.0	28.6	30.7			
CV SD 7.1	n 7	9.2	14.5	12.7	11.5	7.1	6.8	8.6	11.3	10.5	10.0	10.3	12.3	8.9	10.5	10.6	10.0	8.6	9.4	9.6	6.7	9.1	9.0	7.5	8.9	9.6	8.9	14.5	11.4	11.0	7.5	6.8	8.6	10.9	8.7	10.0	9.9	10.9	8.3	10.5	10.1	10.8		
	n 7	8	8	23	8	8	9	25	8	6	8	22	23	22	25	7	6	15	5	10	7	6	13	21	17	38	16	14	8	38	13	13	9	35	15	12	8	35	44	39	25	108		
	PB-SD	M 29.7	23.8	30.1	28.1	24.7	26.5	27.7	26.3	30.1	27.5	26.3	28.1	28.3	25.9	28.1	27.5	29.3	27.4	28.4	33.1	29.7	31.3	31.3	28.4	29.9	31.0	28.5	29.8	29.5	25.6	30.1	28.2	28.5	28.1	27.7	28.2	30.6	28.0	26.3	28.8	29.6	27.2	28.1
PBW SD 9.9	n 9.2	12.7	10.9	8.6	8.0	7.6	8.0	10.5	12.1	6.7	10.0	9.9	9.8	9.4	9.7	9.8	8.7	9.2	10.1	13.0	11.7	10.1	9.7	9.7	9.7	9.7	9.7	9.0	12.7	10.2	10.8	7.6	9.8	10.2	9.6	6.7	9.5	9.9	9.8	9.4	9.8			
	n 18	15	17	10	17	16	17	50	19	15	15	49	54	46	49	149	19	15	34	14	16	30	17	16	33	50	47	97	37	30	107	102	108	31	76	80	36	31	15	82	104	93	49	246
	Gesamt	M 26.3	22.1	27.6	25.5	21.4	25.4	24.0	23.5	21.3	24.5	20.9	22.2	23.0	24.0	24.3	23.7	25.9	24.5	25.3	27.9	26.6	27.2	26.4	23.8	25.1	26.6	25.0	25.8	26.1	23.3	27.6	25.4	24.3	26.0	24.0	24.9	23.7	24.2	20.9	23.4	24.7	24.5	24.3
CV SD 14.7	n 14.8	82	11.6	12.0	10.0	8.8	8.9	9.2	9.1	10.5	10.1	9.8	11.5	9.1	10.4	10.4	11.5	10.0	10.7	10.2	13.2	11.7	10.8	9.5	10.1	10.7	10.9	13.0	9.0	11.6	11.4	10.5	11.1	8.9	10.3	10.1	9.8	10.1	11.2	10.0	10.4	10.6		
	n 14.8	17	50	17	16	17	50	19	15	15	49	149	19	15	34	14	16	30	17	16	30	17	16	33	50	47	97	37	30	107	102	108	31	76	80	36	31	15	82	104	93	49	246	

Subtest 2

[illegible]

Subtest 5

Testreihenfolge	einzel												Präsentation alle												Gesamt																				
	Antwortformat						Antwortformat						Antwortformat						Antwortformat						Antwortformat																				
	einmalig		weiter		korrigierbar		Gesamt		einmalig		weiter		korrigierbar		Gesamt		einmalig		weiter		korrigierbar		Gesamt		einmalig		weiter		korrigierbar		Gesamt														
	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt	ohne	gesamt													
	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe	Zeitvorgabe													
M 203	199	195	199	201	183	216	198	216	189	206	203	206	190	204	200	188	172	178	190	201	195	190	191	191	189	186	188	196	185	195	191	196	190	216	197	202	190	206	198	198	188	204	195		
PRV SD	2,8	3,8	3,7	3,4	3,7	2,1	5,9	4,0	4,3	3,9	4,1	3,5	3,4	4,4	3,8	2,2	3,5	3,0	2,7	3,6	3,1	3,7	2,4	3,1	2,9	3,3	3,1	2,6	3,8	3,7	3,4	3,2	2,8	5,9	3,6	4,0	3,4	3,9	3,7	3,3	4,4	3,6	3,6		
n	9	11	10	30	10	11	7	28	9	9	8	26	28	31	25	84	8	11	9	7	16	10	9	19	27	54	17	22	10	49	18	7	44	19	18	8	3	4,5	5,5	58	25	138			
PC-PB	M 191	194	181	189	177	175	201	182	191	194	175	187	186	187	185	186	170	180	176	171	176	173	170	189	179	170	182	176	181	187	188	174	175	201	179	180	192	175	184	178	185	185	182		
CV SD	2,1	2,4	2,3	3,1	2,2	4,2	3,2	4,9	4,4	4,3	4,4	3,5	3,1	3,6	3,3	2,4	3,8	3,3	1,2	2,5	1,8	2,3	4,0	3,3	1,9	3,5	2,9	2,4	2,3	2,4	2,8	2,3	4,2	2,8	3,8	4,1	4,3	4,0	2,9	3,3	3,6	3,2	3,2		
n	9	11	10	30	10	11	7	28	9	9	8	26	28	31	25	84	8	11	9	7	16	10	9	19	27	54	17	22	10	49	18	7	44	19	18	8	3	4,5	5,5	58	25	138			
PB-PC	M 210	188	199	198	165	201	195	189	204	181	175	186	193	190	189	180	199	189	167	181	175	192	193	193	181	191	186	192	194	199	194	166	191	195	183	198	186	190	175	189	186	190	189	189	
PRV SD	3,3	2,3	2,6	2,4	2,2	1,9	2,6	3,5	2,6	3,1	3,2	3,6	2,4	2,6	2,8	1,9	3,7	3,0	3,1	2,0	2,6	3,1	2,3	2,7	2,8	2,8	2,9	3,1	2,3	2,8	2,7	2,3	2,5	3,1	3,0	3,2	2,6	2,6	2,8	2,6	2,8				
n	5	6	7	18	6	7	8	21	7	8	8	23	18	21	23	62	8	16	6	14	9	6	15	23	22	45	13	14	7	34	12	15	8	35	16	14	8	38	41	43	23	107			
M 218	195	201	204	173	233	164	190	207	186	193	195	199	204	185	196	189	199	194	158	196	180	207	207	188	200	194	200	197	201	199	166	213	164	186	207	195	193	199	193	202	185	195	195		
CV SD	2,4	2,5	2,0	2,4	3,7	2,4	6,2	5,4	3,6	2,7	3,3	3,2	3,7	3,2	4,4	3,9	2,1	2,6	2,4	3,5	2,5	3,4	3,6	1,5	2,9	3,6	2,3	2,0	2,6	2,5	2,0	2,4	3,5	3,0	6,2	4,7	3,5	2,4	3,3	3,1	3,6	2,7	3,5		
n	5	6	7	18	6	7	8	21	7	8	8	23	18	21	23	62	8	16	6	14	9	6	15	23	22	45	13	14	7	34	12	15	8	35	16	14	8	38	41	43	23	107			
M 206	195	196	199	188	190	205	194	211	185	191	195	201	190	197	196	184	183	181	191	186	191	192	191	186	188	187	194	189	196	192	184	190	205	191	200	188	191	194	193	189	197	192	192		
PRV SD	2,9	3,3	3,1	3,1	3,7	2,3	4,2	3,4	3,7	3,6	3,8	3,5	3,0	3,7	3,4	2,0	3,7	3,0	3,0	3,0	3,3	2,2	2,9	2,9	3,1	3,0	2,7	3,5	3,1	3,1	3,3	2,6	4,2	3,3	3,6	3,0	3,8	3,4	3,3	3,0	3,7	3,3	3,3		
n	14	17	17	48	16	18	15	49	16	17	16	49	46	52	48	146	16	19	35	15	15	30	19	15	34	50	49	99	30	36	17	83	31	33	15	79	35	32	16	83	96	101	48	245	
Gesamt	M 201	194	189	194	176	197	181	185	198	191	184	191	191	194	185	190	179	188	184	166	187	176	187	196	191	178	190	184	189	191	189	190	171	192	181	182	192	193	184	191	184	192	185	188	188
CV SD	2,5	2,3	2,4	3,2	3,7	5,6	4,2	4,3	3,6	3,8	3,9	3,6	3,2	4,0	3,6	2,4	3,0	2,4	2,6	2,7	3,4	3,3	3,4	3,3	3,4	2,9	3,1	2,6	3,0	2,4	2,7	2,8	3,2	5,6	3,7	3,9	3,4	3,8	3,7	3,3	3,2	3,4	3,4		
n	14	17	17	48	16	18	15	49	16	17	16	49	46	52	48	146	16	19	35	15	15	30	19	15	34	50	49	99	30	36	17	83	31	33	15	79	35	32	16	83	96	101	48	245	

Subtest 6

Testreihenfolge	Präsentation alle												Antwortformat																															
	einzel						Antwortformat						Gesamt						Antwortformat																									
	weiter			korrigierbar			Gesamt			einmalig			weiter			korrigierbar			Gesamt			einmalig			weiter			korrigierbar			Gesamt													
	Zeitvorgabe			Zeitvorgabe			Zeitvorgabe			Zeitvorgabe			Zeitvorgabe			Zeitvorgabe			Zeitvorgabe			Zeitvorgabe			Zeitvorgabe			Zeitvorgabe			Zeitvorgabe			Zeitvorgabe										
	ohne	gesamt	einzel	ohne	gesamt	einzel	ohne	gesamt	einzel	ohne	gesamt	einzel	ohne	gesamt	einzel	ohne	gesamt	einzel	ohne	gesamt	einzel	ohne	gesamt	einzel	ohne	gesamt	einzel	ohne	gesamt	einzel	ohne	gesamt	einzel											
M 182	219	220	208	173	270	221	221	261	256	236	250	200	249	225	225	270	279	275	231	200	217	231	204	215	242	225	233	221	247	220	231	203	237	221	237	225	228							
PRV SD	5,4	7,3	6,1	6,3	6,2	8,5	9,8	9,0	8,9	7,2	8,8	8,0	7,6	7,7	8,1	8,0	5,0	6,8	5,8	6,7	7,6	7,1	6,1	7,2	6,8	6,2	7,8	7,0	6,8	7,5	6,1	6,9	7,0	8,6	9,8	8,2	7,5	7,5	8,8	7,6	7,2	7,8	8,1	7,6
n	9	11	29	10	10	11	31	7	9	9	25	26	28	31	85	7	8	15	11	9	20	7	10	17	25	27	52	16	17	11	44	21	19	11	51	14	19	51	55	31	137			
PC-PB	M 133	182	178	166	138	189	167	165	200	201	167	188	153	191	171	172	189	198	193	171	177	174	181	172	176	179	181	180	158	189	178	175	155	183	167	168	191	186	167	183	166	186	171	175
CV SD	4,6	5,5	5,0	5,3	5,7	4,2	5,4	5,4	5,5	7,5	5,0	6,1	5,9	5,7	5,0	5,6	4,6	4,9	7,4	6,4	6,8	8,6	6,0	6,9	7,1	5,7	6,3	5,4	6,0	7,0	6,7	5,0	6,4	6,5	5,6	5,0	6,4	6,5	5,6	5,0	5,9			
n	9	9	11	29	10	10	11	31	7	9	9	25	26	28	31	85	7	8	15	11	9	20	7	10	17	25	27	52	16	17	11	44	21	19	11	51	14	19	51	55	31	137		
M 140	254	187	190	233	247	241	240	234	233	198	221	207	243	209	219	213	249	232	220	207	214	208	208	208	213	223	218	181	251	187	211	226	227	241	229	221	219	198	215	210	232	209	218	
PRV SD	8,8	9,0	7,4	9,1	12,0	9,6	7,0	9,2	5,7	6,0	5,7	9,6	7,7	6,7	8,1	5,4	8,1	7,0	7,1	4,8	6,0	3,7	6,9	5,4	5,3	7,0	6,2	7,7	8,1	7,4	8,3	9,3	7,5	7,0	8,0	4,8	6,4	5,7	5,6	7,5	7,3	6,7	7,3	
n	6	5	6	17	7	6	7	20	8	7	8	23	21	18	21	60	9	17	8	6	14	8	9	17	24	48	14	14	6	34	15	12	7	34	16	8	40	45	42	21	108			
PB-PC	M 212	262	188	218	243	228	236	236	228	239	203	222	228	242	210	226	225	277	252	239	200	222	234	220	226	233	236	234	219	271	188	235	241	214	236	230	231	228	203	224	230	239	210	230
CV SD	6,5	5,2	6,9	6,6	10,5	10,2	7,8	9,0	7,4	6,5	6,2	6,6	8,0	7,3	6,9	7,4	7,7	8,3	8,2	7,1	4,2	6,1	5,4	5,2	5,2	6,5	6,9	6,7	7,0	7,1	6,9	7,5	8,5	7,6	7,8	7,9	6,3	5,7	6,2	6,0	7,2	7,0	6,9	7,1
n	6	5	6	17	7	6	7	20	8	7	8	23	21	18	21	60	9	17	8	6	14	8	9	17	24	48	14	14	6	34	15	12	7	34	16	8	40	45	42	21	108			
M 165	231	208	201	198	261	229	229	247	246	218	236	203	247	218	222	239	263	252	226	207	216	219	206	211	228	224	226	202	249	208	222	213	233	222	243	232	218	226	216	235	218	224		
PRV SD	7,0	7,8	6,6	7,4	9,2	8,7	8,6	9,0	7,2	6,6	7,5	7,1	8,4	7,6	7,5	8,0	5,9	7,4	6,7	6,4	6,6	4,9	6,9	6,1	5,9	7,4	6,6	7,4	7,6	7,5	8,1	8,6	8,1	6,2	6,5	6,8	7,5	6,5	7,5	7,5	7,5	7,5		
n	15	14	17	46	17	16	18	51	15	16	17	48	47	46	52	145	15	17	32	19	15	34	15	19	34	49	51	100	30	31	17	78	36	31	18	85	30	35	17	82	96	97	52	245
Gesamt	M 165	211	182	185	181	204	194	193	215	218	184	205	187	211	187	194	208	239	225	199	186	194	209	195	201	205	207	206	186	226	182	201	191	195	194	193	212	205	184	203	196	209	187	199
CV SD	6,6	5,5	6,5	6,3	9,4	7,0	7,1	7,8	6,5	7,1	5,7	6,5	7,8	6,8	6,1	6,9	6,8	7,7	7,4	7,9	5,6	6,9	7,3	6,0	6,5	7,3	6,8	7,0	6,9	7,3	5,5	7,0	8,5	6,3	7,1	7,4	6,8	6,5	7,5	6,8	6,1	7,0		
n	15	14	17	46	17	16	18	51	15	16	17	48	47	46	52	145	15	17	32	19	15	34	15	19	34	49	51	100	30	31	17	78	36	31	18	85	30	35	17	82	96	97	52	245

Subtest 9

Testreihenfolge	Präsentation												Gesamt											
	einzel						alle						einzel						alle					
	Antwortformat						Antwortformat						Antwortformat						Antwortformat					
	einmalig			weiter			einmalig			weiter			einmalig			weiter			einmalig			weiter		
	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe
PC-PB	M 12.1	10.2	11.2	12.1	11.9	12.0	11.1	10.9	11.0	11.6	11.2	11.4	10.4	12.3	11.2	11.7	10.9	12.2	11.4	12.1	10.2	11.2	10.9	12.1
	PBV SD 5.2	3.4	4.4	5.1	3.3	3.8	4.2	5.6	4.5	4.6	3.9	4.2	5.1	5.5	4.3	4.1	5.9	4.4	5.9	5.2	3.4	4.4	5.0	4.4
	n 9	9	18	7	16	23	16	7	23	32	32	64	19	14	33	27	36	24	60	9	18	26	30	56
	CV SD 3.5	5.3	11.8	9.7	9.9	9.8	4.7	4.7	3.4	8.7	8.3	8.5	13.3	12.9	13.1	14.2	15.4	14.6	13.7	14.0	12.3	11.3	11.8	12.3
PB-PC	M 10.0	10.8	10.4	13.2	12.2	12.5	11.6	10.4	11.2	11.5	11.4	10.3	11.4	11.6	11.5	10.6	11.2	10.9	10.0	10.8	10.4	11.2	11.1	11.4
	PBV SD 3.9	2.0	3.1	2.9	4.5	4.0	4.6	3.5	4.2	4.1	3.8	3.9	4.4	3.3	3.8	5.5	3.0	4.8	5.1	3.1	4.3	3.9	4.0	5.0
	n 8	6	14	6	13	19	14	7	21	28	26	54	13	16	29	16	8	24	29	6	14	19	29	48
	CV SD 3.7	4.7	4.0	4.8	3.3	3.8	4.2	4.4	4.2	5.0	4.1	4.6	2.9	3.1	3.0	4.1	2.9	3.7	3.5	3.0	3.3	3.7	4.0	3.5
Gesamt	M 11.1	10.5	10.8	12.6	12.0	12.2	11.3	10.6	11.1	11.5	11.3	11.4	10.1	11.6	10.8	11.7	11.2	11.1	10.5	10.8	10.8	11.4	11.3	11.4
	PBV SD 4.6	2.9	3.8	4.1	3.9	3.9	4.3	4.5	4.3	4.3	3.8	4.0	4.8	4.4	4.6	4.6	4.7	4.6	4.3	4.7	4.1	4.4	4.3	4.4
	n 17	15	32	13	29	42	30	14	44	60	58	118	32	30	62	33	18	51	65	48	11.3	17	15	32
	M 15.0	10.7	13.0	12.1	11.3	11.5	6.7	7.2	6.9	10.2	10.2	10.2	13.6	13.9	13.8	14.3	14.7	14.5	14.0	14.2	14.1	15.0	10.7	13.0
Gesamt	M 15.0	10.7	13.0	12.1	11.3	11.5	6.7	7.2	6.9	10.2	10.2	10.2	13.6	13.9	13.8	14.3	14.7	14.5	14.0	14.2	14.1	15.0	10.7	13.0
	PBV SD 3.5	5.7	5.1	4.6	4.0	4.2	4.3	4.7	4.4	5.5	4.9	5.2	3.2	4.3	3.7	3.1	3.1	3.9	3.5	3.5	5.7	5.1	3.6	4.3
	n 17	15	32	13	29	42	30	14	44	60	58	118	32	30	62	33	18	51	65	48	11.3	17	15	32
	CV SD 3.5	5.7	5.1	4.6	4.0	4.2	4.3	4.7	4.4	5.5	4.9	5.2	3.2	4.3	3.7	3.1	3.1	3.9	3.5	3.5	5.7	5.1	3.6	4.3

Subtest 10

Testreihenfolge	Präsentation												Gesamt											
	einzel						alle						einzel						alle					
	Antwortformat						Antwortformat						Antwortformat						Antwortformat					
	einmalig			weiter			einmalig			weiter			einmalig			weiter			einmalig			weiter		
	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe	ohne	gesamt	Zeitvorgabe
PC-PB	M 35.2	35.9	35.5	31.3	34.5	32.0	31.6	31.8	33.6	33.2	35.9	34.1	34.6	35.5	38.1	37.3	35.7	35.9	35.8	35.2	35.9	35.5	33.4	34.3
	PBV SD 8.0	9.7	8.5	4.6	4.8	4.9	4.7	9.1	7.2	5.9	7.8	7.0	9.1	6.0	6.9	8.4	6.8	8.0	7.6	8.0	9.7	8.5	7.2	5.4
	n 10	8	18	8	16	24	15	17	32	33	41	74	7	18	25	6	15	21	13	33	46	10	8	18
	CV SD 11.3	1.5	10.5	7.4	10.4	9.6	5.7	8.7	7.4	8.4	9.2	8.9	4.8	5.2	5.0	5.7	4.5	4.9	5.1	4.9	4.9	11.3	1.5	10.5
PB-PC	M 30.9	32.0	31.4	35.8	32.3	33.8	29.9	31.7	33.6	31.2	32.3	34.7	33.9	34.2	27.3	30.4	30.7	33.0	32.2	30.9	32.0	31.4	35.3	34.0
	PBV SD 7.2	6.6	6.7	4.6	4.6	6.6	4.8	9.5	7.8	6.7	7.4	7.1	4.1	8.7	7.3	8.7	7.7	6.4	6.9	6.7	6.6	7.1	7.1	6.9
	n 8	8	16	9	12	21	13	15	28	30	35	65	7	13	20	8	15	23	15	28	43	8	16	25
	CV SD 18.5	23.3	27.4	31.7	29.9	28.5	30.6	29.6	28.0	28.2	28.1	21.0	19.6	20.1	15.4	18.7	17.5	18.0	18.5	23.3	24.6	25.4	25.1	23.5
Gesamt	M 33.3	33.9	33.6	33.6	33.5	32.8	30.8	31.8	33.2	32.5	32.8	35.3	34.0	34.4	30.8	35.1	33.7	33.0	34.6	34.1	33.3	33.9	33.6	34.4
	PBV SD 7.7	8.3	7.9	7.1	4.8	5.7	4.7	9.2	7.4	6.3	7.7	7.0	6.8	7.1	7.0	8.0	7.8	8.0	7.6	7.4	7.5	7.7	8.3	6.9
	n 18	16	34	17	28	45	28	32	60	63	76	139	14	31	45	14	30	44	28	61	89	16	34	31
	M 29.7	18.4	24.4	25.6	29.7	28.2	26.8	26.9	26.9	27.3	26.1	18.8	17.9	18.2	15.3	18.3	17.4	17.0	18.1	17.8	29.7	18.4	22.5	23.2
Gesamt	M 29.7	18.4	24.4	25.6	29.7	28.2	26.8	26.9	26.9	27.3	26.1	18.8	17.9	18.2	15.3	18.3	17.4	17.0	18.1	17.8	29.7	18.4	22.5	23.2
	PBV SD 9.4	4.3	9.3	7.0	8.5	8.1	5.7	9.5	7.9	7.3	9.2	8.4	5.1	5.5	6.2	5.6	4.6	5.6	5.4	5.4	9.4	4.3	9.3	7.0
	n 18	16	34	17	28	45	28	32	60	63	76	139	14	31	45	14	30	44	28	61	89	16	34	31
	CV SD 9.4	4.3	9.3	7.0	8.5	8.1	5.7	9.5	7.9	7.3	9.2	8.4	5.1	5.5	6.2	5.6	4.6	5.6	5.4	5.4	9.4	4.3	9.3	7.0

15.8 ANOVA Tabellen der Äquivalenzüberprüfung

In den folgenden Tabellen werden die Ergebnisse der univariaten Varianzanalysen der Ergebnisse der elf Subtests des LPS-neu für die vier Faktoren Präsentation (einzeln, alle), Antwortformat (einmalig, weiter, korrigierbar), Zeitvorgabe (keine, gesamt, einzeln) und Zeitpunkt (CV zu T1 vs. CV zu T2) hinsichtlich des Messwiederholungsfaktors Modus (PBV vs. CV) dargestellt. Angegeben werden die Quadratsummen (QS), die Freiheitsgrade (df ; df_2 entspricht df_{Residual}), die mittleren Quadratsummen ($MQS = QS/df$), die entsprechenden F -Werte und p -Values, das klassische Varianzverhältnis (η_c^2 ; Varianzklärung im Kontext des Gesamtmodells), das partielle Varianzverhältnis (η_p^2 ; Varianzklärung ohne Berücksichtigung der restlichen Faktoren) und die untere und obere Grenze des 80 %-Konfidenzintervalls von η_p^2 .

15.8.1 Subtest 1

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	1100.70	1	1100.70	41.66	<.01	.101	.162	.107	.218
Modus * Reihenfolge	3462.29	1	3462.29	131.03	<.01	.319	.378	.313	.433
Modus * Präsentation	0.57	1	0.57	0.02	.88	<.001	<.001	<.001	.001
Modus * Antwortformat	109.46	2	54.73	2.07	.13	.010	.019	<.001	.042
Modus * Zeitvorgabe	88.66	2	44.33	1.68	.19	.008	.015	<.001	.037
Modus * Reihenfolge * Präsentation	0.14	1	0.14	0.01	.94	<.001	<.001	<.001	.002
Modus * Reihenfolge * Antwortformat	3.43	2	1.72	0.06	.94	<.001	<.001	<.001	.001
Modus * Reihenfolge * Zeitvorgabe	171.01	2	85.51	3.24	.04	.016	.029	.004	.058
Modus * Präsentation * Antwortformat	19.26	2	9.63	0.36	.70	.002	.003	<.001	.012
Modus * Präsentation * Zeitvorgabe	10.92	1	10.92	0.41	.52	.001	.002	<.001	.016
Modus * Antwortformat * Zeitvorgabe	39.88	4	9.97	0.38	.82	.004	.007	<.001	.008
Modus * Reihenfolge * Präs. * Antwortf.	17.90	2	8.95	0.34	.71	.002	.003	<.001	.011
Modus * Reihenfolge * Präs. * Zeitvorgabe	1.21	1	1.21	0.05	.83	<.001	<.001	<.001	.005
Modus * Reihenfolge * Antwortf. * Zeitv.	53.36	4	13.34	0.50	.73	.005	.009	<.001	.012
Modus * Präsentation * Antwortf. * Zeitv.	50.78	2	25.39	0.96	.38	.005	.009	<.001	.025
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	15.59	2	7.80	0.30	.74	.001	.003	<.001	.010
Residualstreuung Messwiederholung	5707.62	216	26.42			.526			
Gesamt-Within-Streuung	10852.79								

15.8.2 Subtest 2

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	131.29	1	131.29	9.81	<.01	.021	.043	.015	.082
Modus * Reihenfolge	2818.43	1	2818.43	210.57	<.01	.450	.494	.434	.542
Modus * Präsentation	7.44	1	7.44	0.56	.46	.001	.003	0	.018
Modus * Antwortformat	28.28	2	14.14	1.06	.35	.005	.010	0	.027
Modus * Zeitvorgabe	35.32	2	17.66	1.32	.27	.006	.012	0	.031
Modus * Reihenfolge * Präsentation	66.58	1	66.58	4.97	.03	.011	.023	.004	.054
Modus * Reihenfolge * Antwortformat	59.71	2	29.85	2.23	.11	.010	.020	0	.045
Modus * Reihenfolge * Zeitvorgabe	26.76	2	13.38	1.00	.37	.004	.009	0	.026
Modus * Präsentation * Antwortformat	4.35	2	2.18	0.16	.85	<.001	.002	0	.004
Modus * Präsentation * Zeitvorgabe	0.60	1	0.60	0.04	.83	<.001	<.001	0	.004
Modus * Antwortformat * Zeitvorgabe	31.70	4	7.92	0.59	.67	.005	.011	0	.016
Modus * Reihenfolge * Präs. * Antwortf.	20.32	2	10.16	0.76	.47	.003	.007	0	.021
Modus * Reihenfolge * Präs. * Zeitvorgabe	27.99	1	27.99	2.09	.15	.004	.010	0	.033
Modus * Reihenfolge * Antwortf. * Zeitv.	35.82	4	8.95	0.67	.61	.006	.012	0	.019
Modus * Präsentation * Antwortf. * Zeitv.	32.23	2	16.11	1.20	.30	.005	.011	0	.029
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	50.26	2	25.13	1.88	.16	.008	.017	0	.040
Residualstreuung Messwiederholung	2891.18	216	13.39			.461			
Gesamt-Within-Streuung	6268.25								

15.8.3 Subtest 3

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	4.79	1	4.79	0.87	.35	.002	.004	0	.022
Modus * Reihenfolge	694.63	1	694.63	125.59	<.01	.328	.369	.304	.424
Modus * Präsentation	7.03	1	7.03	1.27	.26	.003	.006	0	.026
Modus * Antwortformat	1.35	2	0.67	0.12	.89	<.001	.001	0	.001
Modus * Zeitvorgabe	29.49	2	14.74	2.67	.07	.014	.024	.001	.051
Modus * Reihenfolge * Präsentation	1.49	1	1.49	0.27	.60	<.001	.001	0	.014
Modus * Reihenfolge * Antwortformat	7.44	2	3.72	0.67	.51	.004	.006	0	.020
Modus * Reihenfolge * Zeitvorgabe	36.31	2	18.15	3.28	.04	.017	.030	.004	.059
Modus * Präsentation * Antwortformat	35.44	2	17.72	3.20	.04	.017	.029	.004	.058
Modus * Präsentation * Zeitvorgabe	1.04	1	1.04	0.19	.67	<.001	<.001	0	.012
Modus * Antwortformat * Zeitvorgabe	40.31	4	10.08	1.82	.13	.019	.033	0	.053
Modus * Reihenfolge * Präs. * Antwortf.	9.17	2	4.58	0.83	.44	.004	.008	0	.023
Modus * Reihenfolge * Präs. * Zeitvorgabe	0.12	1	0.12	0.02	.88	<.001	<.001	0	.001
Modus * Reihenfolge * Antwortf. * Zeitv.	44.78	4	11.19	2.02	.09	.021	.036	0	.057
Modus * Präsentation * Antwortf. * Zeitv.	5.29	2	2.65	0.48	.62	.002	.004	0	.015
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	12.07	2	6.04	1.09	.34	.006	.010	0	.028
Residualstreuung Messwiederholung	1189.14	215	5.53			.561			
Gesamt-Within-Streuung	2119.87								

15.8.4 Subtest 4

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	1.75	1	1.75	0.50	.48	.002	.002	0	.017
Modus * Reihenfolge	135.23	1	135.23	38.45	<.01	.135	.151	.097	.206
Modus * Präsentation	4.60	1	4.60	1.31	.25	.005	.006	0	.026
Modus * Antwortformat	7.88	2	3.94	1.12	.33	.008	.010	0	.028
Modus * Zeitvorgabe	8.34	2	4.17	1.19	.31	.008	.011	0	.029
Modus * Reihenfolge * Präsentation	0.34	1	0.34	0.10	.76	<.001	<.001	0	.008
Modus * Reihenfolge * Antwortformat	10.51	2	5.25	1.49	.23	.010	.014	0	.034
Modus * Reihenfolge * Zeitvorgabe	22.50	2	11.25	3.20	.04	.022	.029	.004	.057
Modus * Präsentation * Antwortformat	0.14	2	0.07	0.02	.98	<.001	<.001	0	0
Modus * Präsentation * Zeitvorgabe	0.07	1	0.07	0.02	.88	<.001	<.001	0	.001
Modus * Antwortformat * Zeitvorgabe	8.19	4	2.05	0.58	.68	.008	.011	0	.016
Modus * Reihenfolge * Präs. * Antwortf.	18.26	2	9.13	2.60	.08	.018	.023	.001	.049
Modus * Reihenfolge * Präs. * Zeitvorgabe	0.07	1	0.07	0.02	.89	<.001	<.001	0	.001
Modus * Reihenfolge * Antwortf. * Zeitv.	18.54	4	4.63	1.32	.26	.019	.024	0	.039
Modus * Präsentation * Antwortf. * Zeitv.	1.37	2	0.69	0.19	.82	.001	.002	0	.005
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	0.50	2	0.25	0.07	.93	<.001	<.001	0	.002
Residualstreuung Messwiederholung	763.20	217	3.52			.762			
Gesamt-Within-Streuung	1001.49								

15.8.5 Subtest 5

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	21.98	1	21.98	5.03	.03	.017	.023	.004	.054
Modus * Reihenfolge	106.62	1	106.62	24.39	<.01	.084	.102	.057	.153
Modus * Präsentation	1.15	1	1.15	0.26	.61	<.001	.001	0	.013
Modus * Antwortformat	14.89	2	7.44	1.70	.18	.012	.016	0	.037
Modus * Zeitvorgabe	47.11	2	23.56	5.39	<.01	.037	.048	.014	.083
Modus * Reihenfolge * Präsentation	0.22	1	0.22	0.05	.82	<.001	<.001	0	.005
Modus * Reihenfolge * Antwortformat	10.86	2	5.43	1.24	.29	.009	.011	0	.030
Modus * Reihenfolge * Zeitvorgabe	5.50	2	2.75	0.63	.53	.004	.006	0	.019
Modus * Präsentation * Antwortformat	10.34	2	5.17	1.18	.31	.008	.011	0	.029
Modus * Präsentation * Zeitvorgabe	1.14	1	1.14	0.26	.61	<.001	.001	0	.013
Modus * Antwortformat * Zeitvorgabe	17.29	4	4.32	0.99	.41	.014	.018	0	.030
Modus * Reihenfolge * Präs. * Antwortf.	5.63	2	2.81	0.64	.53	.004	.006	0	.019
Modus * Reihenfolge * Präs. * Zeitvorgabe	0.05	1	0.05	0.01	.92	<.001	<.001	0	.002
Modus * Reihenfolge * Antwortf. * Zeitv.	74.84	4	18.71	4.28	<.01	.059	.074	.024	.106
Modus * Präsentation * Antwortf. * Zeitv.	3.10	2	1.55	0.36	.70	.002	.003	0	.012
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	6.21	2	3.11	0.71	.49	.005	.007	0	.020
Residualstreuung Messwiederholung	939.68	215	4.37			.742			
Gesamt-Within-Streuung	1266.61								

15.8.6 Subtest 6

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	545.81	1	545.81	34.63	<.01	.099	.139	.087	.193
Modus * Reihenfolge	1185.00	1	1185.00	75.19	<.01	.215	.259	.196	.318
Modus * Präsentation	0.14	1	0.14	0.01	.92	<.001	<.001	0	.002
Modus * Antwortformat	4.95	2	2.48	0.16	.85	<.001	.001	0	.004
Modus * Zeitvorgabe	16.64	2	8.32	0.53	.59	.003	.005	0	.017
Modus * Reihenfolge * Präsentation	1.51	1	1.51	0.10	.76	<.001	<.001	0	.009
Modus * Reihenfolge * Antwortformat	47.03	2	23.52	1.49	.23	.009	.014	0	.034
Modus * Reihenfolge * Zeitvorgabe	32.15	2	16.07	1.02	.36	.006	.009	0	.026
Modus * Präsentation * Antwortformat	102.06	2	51.03	3.24	.04	.018	.029	.004	.058
Modus * Präsentation * Zeitvorgabe	30.53	1	30.53	1.94	.17	.006	.009	0	.032
Modus * Antwortformat * Zeitvorgabe	30.53	4	7.63	0.48	.75	.006	.009	0	.011
Modus * Reihenfolge * Präs. * Antwortf.	6.10	2	3.05	0.19	.82	.001	.002	0	.006
Modus * Reihenfolge * Präs. * Zeitvorgabe	1.07	1	1.07	0.07	.79	<.001	<.001	0	.007
Modus * Reihenfolge * Antwortf. * Zeitv.	22.86	4	5.72	0.36	.84	.004	.007	0	.007
Modus * Präsentation * Antwortf. * Zeitv.	28.37	2	14.19	0.90	.41	.005	.008	0	.024
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	74.52	2	37.26	2.36	.10	.014	.022	0	.047
Residualstreuung Messwiederholung	3388.53	215	15.76			.614			
Gesamt-Within-Streuung	5517.83								

15.8.7 Subtest 7

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	131.96	1	131.96	16.38	<.01	.053	.073	.034	.120
Modus * Reihenfolge	523.33	1	523.33	64.97	<.01	.212	.238	.175	.297
Modus * Präsentation	7.52	1	7.52	0.93	.34	.003	.004	0	.023
Modus * Antwortformat	11.56	2	5.78	0.72	.49	.005	.007	0	.021
Modus * Zeitvorgabe	8.11	2	4.06	0.50	.61	.003	.005	0	.016
Modus * Reihenfolge * Präsentation	12.47	1	12.47	1.55	.21	.005	.007	0	.030
Modus * Reihenfolge * Antwortformat	4.98	2	2.49	0.31	.73	.002	.003	0	.011
Modus * Reihenfolge * Zeitvorgabe	25.88	2	12.94	1.61	.20	.010	.015	0	.037
Modus * Präsentation * Antwortformat	15.44	2	7.72	0.96	.39	.006	.009	0	.026
Modus * Präsentation * Zeitvorgabe	0.38	1	0.38	0.05	.83	<.001	<.001	0	.005
Modus * Antwortformat * Zeitvorgabe	12.71	4	3.18	0.39	.81	.005	.008	0	.008
Modus * Reihenfolge * Präs. * Antwortf.	1.21	2	0.61	0.08	.93	<.001	<.001	0	.003
Modus * Reihenfolge * Präs. * Zeitvorgabe	1.11	1	1.11	0.14	.71	<.001	<.001	0	.011
Modus * Reihenfolge * Antwortf. * Zeitv.	26.15	4	6.54	0.81	.52	.011	.015	0	.025
Modus * Präsentation * Antwortf. * Zeitv.	3.44	2	1.72	0.21	.81	.001	.002	0	.007
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	10.75	2	5.38	0.67	.51	.004	.006	0	.020
Residualstreuung Messwiederholung	1675.54	208	8.06			.678			
Gesamt-Within-Streuung	2472.55								

15.8.8 Subtest 8

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	218.72	1	218.72	21.35	<.01	.057	.091	.048	.141
Modus * Reihenfolge	1090.49	1	1090.49	106.43	<.01	.283	.333	.268	.391
Modus * Präsentation	27.98	1	27.98	2.73	.10	.007	.013	0	.039
Modus * Antwortformat	39.53	2	19.76	1.93	.15	.010	.018	0	.041
Modus * Zeitvorgabe	61.42	2	30.71	3.00	.05	.016	.027	.003	.056
Modus * Reihenfolge * Präsentation	0.00	1	0.00	0.00	1.00	<.001	<.001	0	0
Modus * Reihenfolge * Antwortformat	9.85	2	4.93	0.48	.62	.003	.004	0	.015
Modus * Reihenfolge * Zeitvorgabe	70.25	2	35.12	3.43	.03	.018	.031	.005	.061
Modus * Präsentation * Antwortformat	2.82	2	1.41	0.14	.87	<.001	.001	0	.003
Modus * Präsentation * Zeitvorgabe	0.01	1	0.01	0.00	.98	<.001	<.001	0	0
Modus * Antwortformat * Zeitvorgabe	9.08	4	2.27	0.22	.93	.002	.004	0	0
Modus * Reihenfolge * Präs. * Antwortf.	18.59	2	9.30	0.91	.41	.005	.008	0	.025
Modus * Reihenfolge * Präs. * Zeitvorgabe	8.66	1	8.66	0.85	.36	.002	.004	0	.022
Modus * Reihenfolge * Antwortf. * Zeitv.	102.42	4	25.60	2.50	.04	.027	.045	.005	.069
Modus * Präsentation * Antwortf. * Zeitv.	1.02	2	0.51	0.05	.95	<.001	<.001	0	0
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	11.40	2	5.70	0.56	.57	.003	.005	0	.017
Residualstreuung Messwiederholung	2182.38	213	10.25			.566			
Gesamt-Within-Streuung	3854.62								

15.8.9 Subtest 9

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	159.36	1	159.36	15.21	<.01	.044	.067	.031	.113
Modus * Reihenfolge	151.71	1	151.71	14.48	<.01	.042	.064	.028	.109
Modus * Präsentation	567.15	1	567.15	54.14	<.01	.156	.204	.144	.263
Modus * Antwortformat	299.21	2	149.60	14.28	<.01	.083	.119	.066	.168
Modus * Zeitvorgabe	27.95	1	27.95	2.67	.10	.008	.012	0	.038
Modus * Reihenfolge * Präsentation	48.09	1	48.09	4.59	.03	.013	.021	.003	.052
Modus * Reihenfolge * Antwortformat	4.83	2	2.41	0.23	.79	.001	.002	0	.007
Modus * Reihenfolge * Zeitvorgabe	4.23	1	4.23	0.40	.53	.001	.002	0	.016
Modus * Präsentation * Antwortformat	59.77	1	59.77	5.71	.02	.016	.026	.006	.060
Modus * Präsentation * Zeitvorgabe	8.18	1	8.18	0.78	.38	.002	.004	0	.021
Modus * Antwortformat * Zeitvorgabe	46.70	2	23.35	2.23	.11	.013	.021	0	.046
Modus * Reihenfolge * Präs. * Antwortf.	21.25	1	21.25	2.03	.16	.006	.010	0	.033
Modus * Reihenfolge * Präs. * Zeitvorgabe	0.05	1	0.05	0.00	.94	<.001	<.001	0	0
Modus * Reihenfolge * Antwortf. * Zeitv.	5.68	2	2.84	0.27	.76	.002	.003	0	.009
Modus * Präsentation * Antwortf. * Zeitv.	0.03	1	0.03	0.00	.96	<.001	<.001	0	0
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	10.90	1	10.90	1.04	.31	.003	.005	0	.024
Residualstreuung Messwiederholung	2210.40	211	10.48			.610			
Gesamt-Within-Streuung	3625.48								

15.8.10 Subtest 10

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	11739.76	1	11739.76	433.02	<.01	.538	.676	.631	.709
Modus * Reihenfolge	452.11	1	452.11	16.68	<.01	.021	.074	.035	.121
Modus * Präsentation	2610.69	1	2610.69	96.29	<.01	.120	.316	.251	.375
Modus * Antwortformat	232.91	2	116.46	4.30	.01	.011	.040	.009	.073
Modus * Zeitvorgabe	101.04	1	101.04	3.73	.05	.005	.018	.002	.047
Modus * Reihenfolge * Präsentation	25.42	1	25.42	0.94	.33	.001	.004	0	.023
Modus * Reihenfolge * Antwortformat	47.54	2	23.77	0.88	.42	.002	.008	0	.025
Modus * Reihenfolge * Zeitvorgabe	16.15	1	16.15	0.60	.44	<.001	.003	0	.019
Modus * Präsentation * Antwortformat	7.26	1	7.26	0.27	.61	<.001	.001	0	.014
Modus * Präsentation * Zeitvorgabe	65.84	1	65.84	2.43	.12	.003	.012	0	.037
Modus * Antwortformat * Zeitvorgabe	765.92	2	382.96	14.13	<.01	.035	.120	.066	.169
Modus * Reihenfolge * Präs. * Antwortf.	0.72	1	0.72	0.03	.87	<.001	<.001	0	.003
Modus * Reihenfolge * Präs. * Zeitvorgabe	105.93	1	105.93	3.91	.05	.005	.018	.002	.048
Modus * Reihenfolge * Antwortf. * Zeitv.	12.85	2	6.42	0.24	.79	<.001	.002	0	.008
Modus * Präsentation * Antwortf. * Zeitv.	1.31	1	1.31	0.05	.83	<.001	<.001	0	.005
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	0.06	1	0.06	0.00	.96	<.001	<.001	0	0
Residualstreuung Messwiederholung	5639.19	208	27.11			.258			
Gesamt-Within-Streuung	21824.70								

15.8.11 Subtest 11

Faktor	QS	df	MQS	F	p	η_c^2	η_p^2	$u_{\eta_p^2}$	$o_{\eta_p^2}$
Modus	150.38	1	150.38	9.50	<.01	.038	.048	.016	.091
Modus * Reihenfolge	475.47	1	475.47	30.05	<.01	.121	.137	.082	.194
Modus * Präsentation	9.27	1	9.27	0.59	.44	.002	.003	0	.021
Modus * Antwortformat	36.95	2	18.48	1.17	.31	.009	.012	0	.033
Modus * Zeitvorgabe	46.30	2	23.15	1.46	.23	.012	.015	0	.038
Modus * Reihenfolge * Präsentation	0.00	1	0.00	0.00	1.00	<.001	<.001	0	0
Modus * Reihenfolge * Antwortformat	10.37	2	5.19	0.33	.72	.003	.003	0	.012
Modus * Reihenfolge * Zeitvorgabe	14.01	2	7.00	0.44	.64	.004	.005	0	.016
Modus * Präsentation * Antwortformat	0.71	2	0.35	0.02	.98	<.001	<.001	0	0
Modus * Präsentation * Zeitvorgabe	8.39	1	8.39	0.53	.47	.002	.003	0	.020
Modus * Antwortformat * Zeitvorgabe	52.57	4	13.14	0.83	.51	.013	.017	0	.028
Modus * Reihenfolge * Präs. * Antwortf.	41.52	2	20.76	1.31	.27	.011	.014	0	.035
Modus * Reihenfolge * Präs. * Zeitvorgabe	1.92	1	1.92	0.12	.73	<.001	<.001	0	.011
Modus * Reihenfolge * Antwortf. * Zeitv.	56.68	4	14.17	0.90	.47	.014	.019	0	.030
Modus * Präsentation * Antwortf. * Zeitv.	10.69	2	5.35	0.34	.71	.003	.004	0	.013
Modus * Reihenf. * Präs. * Antwortf. * Zeitv.	2.70	2	1.35	0.09	.92	<.001	<.001	0	.004
Residualstreuung Messwiederholung	3006.78	190	15.83			.766			
Gesamt-Within-Streuung	3924.71								